МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
УО "БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ"

Кафедра лесоустройства

ЛЕСНАЯ БИОМЕТРИЯ

Конспект лекций по одноименной дисциплине для студентов спец. 1-75 01 01 «Лесное хозяйство»

Составитель к. с.-х. н., доцент В.П. Машковский

Минск 2010

ВВЕДЕНИЕ

Лес – это сложная биологическая система, состоящая из множества элементов, которые можно разделить на однородные группы. С одной стороны, элементы таких групп сходны между собой, так как все они имеют ряд одинаковых свойств. Например, все деревья в лесу имеют ствол, ветви и крону. Деревья одной породы имеют одинаковую кору, форму листьев или хвои, одинаковую структуру кроны, форму ствола и т. д. Вместе с тем каждый элемент индивидуален и отличается от других элементов группы. Такие отличия, как правило, носят количественный характер. Например, деревья могут отличаться друг от друга по диаметру на высоте груди, высоте, объему ствола и т. д.

Довольно легко выделить в той или иной степени однородные группы по различным качественным признакам во время анализа лесных сообществ. Например, не составит труда в сосново-березовом лесу выделить отдельные совокупности деревьев березы или сосны. Как правило, следующий шаг анализа природных объектов предполагает получение каких-либо количественных показателей, описывающих однородные совокупности. Однако как получить такие характеристики, если все элементы совокупности имеют разные значения количественных признаков? Например, при получении информации о высоте березовых деревьев в сосново-березовом лесу встает вопрос, какое дерево выбрать в качестве эталона, если высоты у всех разные? Или, если измерять все березы в древостое, как потом разобраться в этом огромном количестве данных? Такого рода проблемы постоянно встают в ходе исследования природных объектов.

Для того чтобы проанализировать большое количество данных, необходимо иметь совокупность методов, позволяющих обрабатывать массовые наблюдения и делать обоснованные выводы. Потребность в таких методах появилась еще в XVII веке и была вызвана накоплением огромного количества сведений в области страхового дела, демографии, здравоохранения, торговли и т. д. Это привело к возникновению математической статистики, которая развивалась параллельно с теорией вероятностей и широко использовала математический аппарат, предоставляемый последней.

Большой вклад в становление этого направления в науке внес бельгийский антрополог Кетле (1796 – 1874). Он показал, что самые различные физические особенности человека, и даже его поведение, подчиняются определенным статистическим закономерностям. Позднее Кетле пришел к заключению, что этим статистическим закономерностям подчиняется не только человеческое общество, но и все другие живые существа.

В биологии использование ряда математических приемов и методов началось практически только в XIX веке. Ф. Гальтон (1822 – 1911) разработал основы новой науки, для которой в 1889 году предложил новый термин «биометрия». Позднее, в 1899 году, Г. Дункер предложил другое название – «вариационная статистика».

В XIX веке английские ученые во главе с Ф. Гальтоном и К. Пирсоном (1857 – 1936) создали математический аппарат биометрии. Ученик К. Пирсона В. Госсет (1876 – 1937), известный под псевдонимом «Стьюдент», положил начало теории малой выборки. Наряду с ним в дальнейшем развитии этой теории принимал активное участие Р. Фишер (1890 – 1962). Последний также заложил основы теории планирования экспериментов.

В русском лесном хозяйстве биометрические методы начали применяться в начале XX века. Впервые в СССР их использование в лесном деле было описано в 1929 году. в книге Голубева В. В. «Элементы математической статистики в приложении к лесному делу» [6]. В настоящее время развитие вычислительной техники в значительной степени облегчает выполнение различных расчетов, что является мощным толчком к широкому использованию в решении лесохозяйственных задач классических биометрических методов, а также к расширению перечня применяемых методик.

1. ОСНОВНЫЕ ПОНЯТИЯ

1.1. Предмет и задачи лесной биометрии

Лесная биометрия в решении своих задач использует методы, разработанные математической статистикой и опирающиеся на основные понятия теории вероятностей. Приведенный математический аппарат создан для обработки массовых (статистических) данных. Экспериментальные наблюдения, собранные в лесу, как раз и являются такими данными. Они представляют собой числовые характеристики массовых случайных явлений (высота, диаметр или объем дерева; число шишек на дереве; количество семян в шишке и т. д.). В связи с этим предметом лесной биометрии являются различные случайные величины, характеризующие лесные сообщества.

Основная задача лесной биометрии – это количественный и качественный анализ массовых случайных явлений, имеющих место в лесных биогеоценозах. Такая процедура подразумевает ряд этапов: организация и планирование наблюдений; сбор экспериментальных данных; свертка информации, т. е. сведение большого количества исходных данных к небольшому числу параметров, которые в сжатом виде характеризуют всю исследуемую совокупность; анализ экспериментальных данных; принятие решений на основе результатов такого анализа; прогнозирование случайных явлений.

Анализ экспериментальных данных подразумевает проверку гипотез о модели распределения (в частности, о функции распределения $F(x)$) случайной величины; проверку гипотез о параметрах этих распределений; получение точечных или интервальных оценок для данных параметров.

1.2. Особенности лесохозяйственной информации

При анализе лесных сообществ экспериментальные данные собираются путем регистрации значений каких-либо признаков (свойств), позволяющих отличить один объект изучаемой совокупности от другого. Как правило, такие показатели характеризуются изменчивостью своих значений при переходе от наблюдения к наблюдению (от объекта к объекту). Это явление называется *вариацией* (от лат. variatio – изменение колебания), а отдельные числовые значения меняющегося параметра – *вариантами* (от лат. varians, variantis – различимый, изменяющийся), или *датами*. Вариация вызывается естественной изменчивостью и ошибками измерений. Так как случайные ошибки обычно невелики в сравнении с естественным варьированием, то колебания результатов наблюдений рассматриваются обычно как естественное варьирование признаков. Вариация может быть *качественной*, *или атрибутивной*, если различия между вариантами выражаются в каких-то качествах. В этом случае признаки не поддаются непосредственному измерению. Как пример качественного варьирования можно привести состояние куколок соснового шелкопряда. Данный признак для каждой учтенной куколки мог принимать следующие значения: паразитированная, погибшая по другим причинам или здоровая [24].

Если качественный признак может принимать только два значения, то такая вариация называется *альтернативной*, а признак – *альтернативным*. Например, при определении плотности популяции соснового шелкопряда возраст учитываемых гусениц мог принимать значения: гусеница младших возрастов или гусеница старших возрастов [24].

Очень часто изучаемые свойства поддаются количественной оценке. Это означает, что имеет место *количественная вариация*. В таком случае признаки могут быть *мерными, или метрическими* (например, высота или диаметр дерева). В данном случае имеет место *непрерывная вариация*. Это означает, что величина признаков может принимать любые числовые значения в каком-либо интервале. Кроме мерных (метрических), существуют *счетные*, *или меристические*, признаки. Для них характерна *дискретная* (*прерывная*) *вариация*. В этом случае значения таких признаков не могут отличаться друг от друга менее чем на некоторую конечную величину. Как правило, значения таких признаков выражаются целыми числами. В качестве примера можно привести количество шишек на деревьях сосны (данный признак использовался для характеристики различия деревьев по плодоношению) [17].

Как уже отмечалось выше, анализ массовых случайных явлений, имеющих место в лесных биогеоценозах, начинается со сбора экспериментальных данных. На указанном этапе, как правило, выполняются измерения различных параметров, характеризующих изучаемые объекты. Измерения проводятся с помощью различных приборов и инструментов, которые обеспечивают получение результатов с определенной точностью. Для того чтобы правильно выбрать способ проведения измерений, подобрать инструменты, обеспечивающие необходимую точность полученных результатов, исследователь должен обладать общими сведениями об измерениях. Познакомимся с основными понятиями, касающимися измерений.

Результаты любых измерений не являются абсолютно точными. Они содержат погрешность, или ошибку.

Погрешностью измерения называется отклонение результата измерения от истинного значения измеряемой величины.

В свою очередь, *результат измерения* – это оценка измеряемой физической величины в виде некоторого числа принятых для нее единиц, полученная путем измерения, а *истинное значение физической величины* – значение физической величины, которое идеальным образом отражает как в качественном, так и в количественном отношении соответствующее свойство объекта.

По форме выражения различают абсолютные и относительные погрешности.

Абсолютная погрешность – погрешность, выраженная в единицах измеряемой величины и представляющая разность между полученным результатом измерения (x_i) и истинным значением измеряемой величины (X):

$$\Delta = x_i - X .$$

Относительная погрешность – погрешность, выраженная в долях истинного значения измеряемой величины:

$$\delta = \frac{x_i - X}{X} .$$

Относительные погрешности чаще всего выражаются в процентах.

Наряду с погрешностью измерения характеризуются точностью. *Точность измерений* – это качество измерений, отражающее близость их результатов к истинному значению измеряемой величины.

Погрешности в измеряемых величинах могут быть вызваны разными причинами. Основные источники погрешностей – это изменение условий наблюдений (влажности, температуры, давления и т. д.), несовершенство применяемого метода измерений и субъективизм наблюдения.

По способу получения числового значения измеряемой величины все измерения делятся на четыре основных вида: прямые, косвенные, совокупные и совместные.

Прямыми называют измерения, заключающиеся в экспериментальном сравнении измеряемой величины с мерой этой величины, или в отсчете показаний средства измерений, непосредственно дающего значение измеряемой величины (например, измерение длины линейкой).

Косвенные – это такие измерения, результат которых определяют на основании прямых измерений величин, связанных с измеряемой величиной определенной зависимостью (пример – измерение объема параллелепипеда по длине, ширине и высоте).

Совокупными называют измерения, в которых значение измеряемых величин определяют по данным повторных измерений одной или нескольких одноименных величин при различных сочетаниях мер или самих этих величин. Результаты совокупных измерений находят путем решения системы уравнений, составляемых по результатам нескольких прямых измерений.

Совместными называют производимые одновременно (прямые или косвенные) измерения двух или нескольких неодноименных величин. Цель совместных измерений, по существу, – это нахождение функциональной зависимости между величинами.

По точности оценивания погрешностей измерения можно классифицировать следующим образом:

– *измерения с точным оцениванием погрешностей*, при которых учитывают индивидуальные свойства средств измерений и контролируют условия;

– *измерения с приближенным оцениванием погрешностей*, при которых учитывают нормативные данные о свойствах средств измерений и приближенно оценивают условия. Наиболее существенные влияющие величины иногда измеряют;

– *измерения с предварительным оцениванием погрешностей*, при которых регламентированы типы (марки) применяемых средств измерений, условия выполнения измерений и заранее оценены погрешности. Измерения с предварительным оцениванием погрешностей часто называют *техническими* измерениями.

По закономерности появления погрешности измерений могут быть систематическими и случайными.

Систематической погрешностью измерения называется составляющая погрешности измерения, которая остается постоянной или закономерно изменяется при повторных измерениях одной и той же величины.

Различают постоянные и закономерно изменяющиеся систематические погрешности. В свою очередь, закономерно изменяющиеся систематические погрешности бывают прогрессирующие, периодические и изменяющиеся по сложному закону.

Постоянная систематическая погрешность – это погрешность, остающаяся неизменной и потому повторяющаяся при каждом наблюдении или измерении.

Прогрессирующие погрешности – погрешности, постоянно возрастающие или убывающие при измерениях.

Периодические погрешности – погрешности, меняющиеся с определенным периодом.

В общем случае систематическая погрешность может меняться по сложному непериодическому закону.

Обнаруженная и оцененная систематическая погрешность исключается из результата измерения путем введения поправки. Однако таким образом полностью устранить систематическую погрешность не представляется возможным.

Систематическую погрешность измерения можно обнаружить либо сопоставлением данного результата с результатом измерения этой же величины, но полученным другим методом, либо путем использования более точных средств измерений.

Однако обычно систематические погрешности оценивают путем теоретического анализа условий измерения, основываясь на известных свойствах средств измерения.

Качество измерений, отражающее близость к нулю систематических погрешностей результатов измерений, называют *правильностью измерений*.

Говоря о свойствах погрешностей, будем отличать также грубые погрешности, или промахи.

Грубой погрешностью называют такую, которая, существенно превышает погрешность, оправданную условиями и методом измерения, свойствами примененных средств измерений, квалификацией экспериментатора.

Грубые погрешности при статистических измерениях обнаруживают статистическими методами и обычно исключают из рассмотрения.

Промахи – следствие неправильных действий экспериментатора (описка, неправильное снятие показания прибора и т. д.). Промахи обнаруживают нестатистическими методами, и их следует всегда исключать из рассмотрения.

Погрешности измерений делят также на *статические* и *динамические*.

Динамические погрешности – это погрешности, обусловленные инерционными свойствами средств измерений.

Методические погрешности – те, которые могут возникать вследствие недостаточной разработанности теории явлений, положенных в основу измерения, и неточности тех соотношений, которые используются для нахождения оценки измеряемой величины.

Инструментальные погрешности – погрешности из-за несовершенства средств измерений. Обычно различают *основную погрешность* средств измерений – погрешность в условиях, принятых за нормальные, и *дополнительные*, т. е. погрешности, обусловленные отклонением влияющих величин от их нормальных значений.

Личные погрешности. Индивидуальные особенности лица, выполняющего измерения, обусловливают появление индивидуальных, свойственных данному лицу погрешностей (неправильный отсчет десятых долей шкалы прибора и т. д.).

Для оценивания погрешностей измерений необходимы сведения о погрешностях средств измерений. Такая информация формируется в процессе нормирования погрешностей средств измерений. Погрешность средств измерений является их специфической метрологической характеристикой.

С целью оценки совокупного влияния разных погрешностей на точность измерений устанавливается класс точности средств измерений.

Класс точности – это обобщенная характеристика средств измерений, определяющая допустимые пределы для всех погрешностей этих средств измерений, а также и все другие свойства средств измерений, влияющие на их точность.

Погрешности средств измерений могут быть выражены в виде *абсолютных* или *относительных* величин.

Одним из способов увеличения точности измерения является обнаружение и устранение систематических погрешностей. Обнаружение систематических погрешностей представляет собой сложную задачу. Особенно трудно находить постоянную систематическую погрешность. Для ее обнаружения целесообразно выполнить измерение несколькими (хотя бы двумя) различными путями или параллельно выполнять измерения с применением более точных средств измерения. Изменяющиеся систематические погрешности, кроме того, можно обнаружить с помощью различных статистических методов.

Если систематическая погрешность найдена, то обычно удается ее оценить и устранить. Устранение систематических погрешностей часто осуществляется путем математической обработки опытных данных, применением соответствующих методов измерений.

Различают еще ряд методов устранения систематических погрешностей.

Метод замещения заключается в замене измеряемой величины известной величиной так, что при этом в состоянии и действии всех используемых средств измерений не происходит никаких изменений (при взвешивании предмет замещают гирями, не выводя весы из равновесия, и оценивают вес гирь).

Метод противопоставления – измерение выполняется с двумя наблюдениями, проводимыми так, чтобы возникновение постоянной погрешности оказывало разные, но известные по закономерности воздействия на результаты наблюдений.

Рассмотрим пример. При взвешивании предмет кладут сначала на одну чашку весов, а на вторую – гири весом p_1, уравновешивая весы. Тогда вес x можно определить следующим образом:

$$x = \frac{l_2}{l_1} \cdot p_1, \tag{1}$$

где l_1 и l_2 – длина плеч весов.

Потом предмет перекладывают на другую чашку весов, вновь уравновешивая его гирями весом p_2. В этом случае вес предмета можно определить по формуле

$$x = \frac{l_1}{l_2} \cdot p_2. \tag{2}$$

Если теперь умножить уравнение (1) на уравнение (2), получим

$$x \cdot x = \frac{l_2}{l_1} \cdot p_1 \cdot \frac{l_1}{l_2} \cdot p_2 = p_1 \cdot p_2.$$

Выразив из последнего выражения x, будем иметь следующую формулу для вычисления веса предмета по результатам двух взвешиваний:

$$x = \sqrt{p_1 \cdot p_2}\,. \qquad (3)$$

Полученное выражение не зависит от длин плеч весов и, следовательно, погрешность весов, вызванная разной длиной плеч, не будет влиять на полученный результат.

Метод компенсации погрешности по знаку предусматривает измерение с двумя наблюдениями, выполняемыми так, чтобы постоянная систематическая погрешность в результате каждого из них входила с разными знаками. Например, при измерении румбов с помощью буссоли для того, чтобы компенсировать погрешность, возникающую из-за несовпадения оси магнитной стрелки с центром кольца градусных делений, отсчеты берут по двум противоположным концам магнитной стрелки. Среднее арифметическое значение из этих двух отсчетов будет свободным от погрешности за счет эксцентриситета.

Если измерение какой-то величины выполнено несколько раз и между результатами отдельных измерений имеются различия, которые индивидуально непредсказуемы, а какие-либо присущие им закономерности проявляются лишь на значительном числе данных, то погрешность, обусловленную таким рассеиванием результатов измерений, называют *случайной*.

Случайные погрешности обнаруживаются путем повторения измерения одной и той же величины в одних и тех же условиях.

Закономерность проявления случайных погрешностей поддается учету при достаточно большом числе измерений.

Случайные погрешности подчиняются закономерности статистического характера. Это обусловливает наличие некоторых свойств у случайных погрешностей.

1. Случайные погрешности не могут превосходить по абсолютному значению определенного предела. Этот предел зависит от условий, в которых производятся измерения (средства измерения, оператор, внешние условия и др.).
2. Случайные погрешности – положительные и отрицательные – одинаково часто встречаются в ряду измерений.
3. Среднее арифметическое из случайных погрешностей измерений одной и той же величины, произведенных в одинаковых условиях, стремится к нулю при неограниченном возрастании числа измерений. Это очень важное свойство компенсации случайных погрешностей, проявляющееся при их сложении, является следствием первых двух свойств случайных

погрешностей.

4. Чем больше абсолютное значение погрешности, тем реже такая погрешность встречается в ряду измерений. Данное свойство присуще абсолютному большинству рядов случайных погрешностей и имеет большое значение в практической деятельности, так как дает возможность определить практический предел, который не должны превосходить случайные погрешности в конкретных условиях измерений.

Когда погрешности измерений обладают вышеперечисленными свойствами, то говорят о нормальном распределении погрешностей измерений.

Качество измерений, отражающее близость получаемых результатов при выполнении работ в одних и тех же условиях, называются *сходимостью результатов измерений*. Хорошая сходимость свидетельствует о малости случайных погрешностей.

Знания об измерениях, о возможных ошибках, о точности приборов и инструментов позволят исследователю в ходе сбора экспериментального материала минимизировать ошибки и получить правильные и достоверные данные.

1.3. Выборочный метод. Генеральная и выборочная совокупности

Любые статистические исследования используют в качестве исходной информации совокупность результатов каких-либо экспериментов или наблюдений над объектами, представленных в большинстве случаев в виде некоторых числовых данных. Наблюдения могут быть *сплошными* (*полными*), когда охватываются все члены изучаемой совокупности, или *частичными* (*выборочными*), когда обследуется лишь часть членов данной совокупности. Сплошные наблюдения позволяют получить наиболее точную и исчерпывающую информацию об изучаемом объекте, но зачастую бывают очень трудоемкими или вообще невозможными. На практике чаще анализируется некоторая часть исследуемой совокупности и на основании полученных результатов дается приближенная оценка свойств всего группового объекта в целом. Например, для 45-летнего соснового древостоя площадью 5,3 га необходимо собрать сведения о высотах деревьев. Число стволов в таком насаждении составит приблизительно 6–7 тыс. штук. Измерение такого количества высот – весьма трудоемкая задача. Вместо сплошного обследования в древостое закладывают так называемые пробные площади (отграниченные каким-либо образом небольшие участки леса) и выполняют измерения только для попавших туда деревьев. Как правило, число последних бывает около двухсот штук. Вся совокупность деревьев в исследуемом древостое называется *генеральной совокупностью*, а та часть, которую подвергли измерениям, – *выборочной совокупностью*. В общем случае под *генеральной совокупностью* понимают такую, которая содержит в себе все элементы изучаемого группового объекта. Число таких элементов называется объемом генеральной совокупности. *Выборочной совокупностью* (*выборкой*) объема n называется совокупность n объектов, отобранных из исследуемой генеральной совокупности. Метод, заключающийся в том, что по материалам выборки делают заключение о свойствах генеральной совокупности, называется *выборочным методом*.

Для того чтобы по результатам выборки получать достаточно точные характеристики генеральной совокупности, выборка должна быть *представительной*, *или репрезентативной* (от лат. represento – представляю), т. е. достаточно хорошо отражать свойства исследуемой случайной величины. Репрезентативной выборка является в том случае, если вероятность попадания в нее для любого из элементов генеральной совокупности одинакова. Это достигается рандомизацией (от англ. random – случай), которая заключается в том, что отбор элементов из генеральной совокупности должен быть случайным.

Есть несколько способов отбора объектов из генеральной совокупности.

Случайная выборка с возвращением (*повторный отбор*). В этом случае из генеральной совокупности случайным образом отбирается объект, анализируется и снова возвращается в генеральную совокупность. В связи с этим любой объект может попасть в выборку повторно.

Случайная выборка без возвращения (*бесповторный отбор*). В этом случае отобранные случайным образом из генеральной совокупности элементы после анализа не возвращаются обратно и поэтому они не могут повторно попасть в выборку. В лесном хозяйстве, как правило, используют бесповторный отбор. Например, когда для получения информации о древостое в нем отводится пробная площадь, а потом анализируются все деревья, произрастающие на ней, то каждое отдельное дерево из исследуемого насаждения может быть учтено только один раз.

Типическая (*стратифицированная*) *выборка.* В этой выборке предварительно генеральная совокупность делится на отдельные типические группы (страты), обладающие внутренней однородностью и различающиеся между собой. Затем из каждой группы случайным образом отбирают объекты для анализа. Число элементов, отбираемых для анализа из каждой группы, как правило, бывает пропорционально объему группы.

Серийная (*гнездовая*) *выборка.* В этом случае генеральная совокупность, так же как и при типической выборке, разбивается на части. Однако, в отличие от типической выборки, группы выделяются, как правило, по территориальному признаку. Затем из этих частей (гнезд) каким-либо образом выбирают некоторые и производят совместный анализ всех элементов, содержащихся в отобранных гнездах. Другими словами, при данной выборке отбираются не отдельные элементы, а целые группы элементов. Гнезда при серийном отборе могут формироваться как одинакового размера, так и содержащие разное число элементов.

Механическая выборка. В данном случае генеральная совокупность механически разбивается на отдельные группы, как правило, равной величины, и затем из каждой группы для анализа отбирается случайным образом по одному объекту. При этой выборке число образованных групп соответствует размеру выборки. Часто механическая выборка осуществляется путем отбора объектов через равный интервал (например, каждый десятый, сотый объект и т. д.). Такой вариант механического отбора называют *систематической выборкой*.

1.4. Группировка первичных данных.

Массовый экспериментальный материал, полученный в результате проведения каких-либо опытов или в результате учета (измерения) различных признаков исследуемых объектов, обычно заносится в учетные карточки или рабочие журналы. Как уже отмечалось выше, регистрируемые параметры исследуемого объекта, как правило, изменяются от опыта к опыту (варьируют). Для того чтобы проанализировать такие данные, их лучше всего каким-либо образом предварительно сгруппировать. Как известно, характер варьирования может быть разным. В зависимости от вида изменчивости применяются различные способы группировки первичного статистического материала. В случае качественной вариации исходные данные лучше всего группировать по значениям исследуемого признака. Например, при анализе коконов соснового шелкопряда в центре очага, выявленного в 1966 г. в Новобобовичском стационаре, куколки подразделялись на погибшие от заражения паразитами, погибшие по другим причинам и здоровые [24]. В данном случае мы имеем дело с качественной вариацией признака. Для проводимого примера группировка будет заключаться в подсчете числа куколок, погибших от заражения паразитами, погибших по другим причинам и здоровых (табл. 1).

Таблица 1. **Смертность соснового шелкопряда в фазе куколки**

Типы куколок	Количество куколок	Процент от общего количества
Паразитированные куколки	1307	98
Погибшие по другим причинам	13	1
Здоровые	15	1
Итого	1335	100

Наиболее простым случаем качественной вариации является альтернативная вариация, когда признак может принимать только два значения. Например, учитываемые на пробных площадях деревья могут подразделяться на семеносящие и несеменосящие (табл. 2). Приводимый пример взят из работы, посвященной исследованию влияния атмосферных промышленных загрязнений на семеношение сосны [13].

Таблица 2. **Учет семеношения на пробных площадях**

Номер пробной площади	Число деревьев, шт.		
	семеносящих	несеменосящих	всего
1	39	268	307
2	6	375	381
3	110	298	408

Если для исследуемого признака характерна количественная вариация, т. е. исследуемый признак принимает различные числовые значения, то в этом случае используются другие способы группировки исходных данных. Разберем случай, когда признак характеризуется количественной дискретной вариацией. В качестве примера такой вариации можно привести данные о числе порослевин от основания усохших после пожара берез [11]. В первой строке табл. 3 записаны номера наблюдений, а во второй – число порослевин для каждого наблюдения (дерева).

Таблица 3. **Характеристика возобновления березы**

Номер ствола	Число порослевин	Номер ствола	Число порослевин
1	17	11	5
2	8	12	14
3	13	13	17
4	20	14	3
5	7	15	2
6	5	16	2
7	31	17	3
8	16	18	4
9	3	19	5
10	10	20	20

Такая запись данных называется простым статистическим рядом. Если экспериментальные данные упорядочить по возрастанию признака и перенумеровать, то получим вариационный ряд.

В табл. 4 приведен вариационный ряд, составленный по материалам анализа величины сентябрьских выводков сеголеток рыжей полевки в Беловежской пуще [4].

Таблица 4. **Величина сентябрьских выводков у сеголеток (первые две генерации) рыжей полевки**

№ п/п	Число эмбрионов	№ п/п	Число эмбрионов	№ п/п	Число эмбрионов	№ п/п	Число эмбрионов
1	2	12	4	23	5	34	6
2	2	13	4	24	5	35	6
3	2	14	4	25	5	36	6
4	2	15	4	26	5	37	6
5	3	16	4	27	5	38	6
6	3	17	4	28	5	39	6
7	3	18	4	29	5	40	7
8	3	19	4	30	5	41	7
9	3	20	5	31	6	42	7
10	3	21	5	32	6	43	7

№ п/п	Число эмбрионов	№ п/п	Число эмбрионов	№ п/п	Число эмбрионов	№ п/п	Число эмбрионов
11	3	22	5	33	6	44	8

В общем виде вариационный ряд можно записать следующим образом:

$$x_1^{(n)} \le x_2^{(n)} \le \ldots \le x_n^{(n)}.$$

Члены вариационного ряда $x_m^{(n)} \left(m = 1, 2, \ldots, n\right)$ называются порядковыми, или ранговыми, статистиками. Отношение *m/n* называется рангом члена ряда $x_m^{(n)}$.

В последнем приводимом примере есть наблюдения, имеющие одинаковые значения. Это позволяет произвести группировку данных (табл. 5).

Таблица 5. **Распределение сеголеток (первые две генерации) рыжей полевки по величине сентябрьских выводков**

Классы (число эмбрионов)	Частоты (количество самок)	Частости (доля самок в каждом классе)	Накопленные частоты
2	4	0,09	4
3	7	0,16	11
4	8	0,18	19
5	11	0,25	30
6	9	0,21	39
7	4	0,09	43
8	1	0,02	44
Итого	44	1,00	

В полученном сгруппированном статистическом ряду (первая колонка таблицы) в порядке возрастания перечислены все встречающиеся значения исследуемого признака. Это классы сгруппированного статистического ряда. Во второй колонке записаны количества самок в выборке, имеющих соответствующее классу число эмбрионов. Эти величины называются *частотами*. Иногда вместо них используются относительные величины, называемые *относительными частотами*, или *частостями* (3 колонка табл. 5). Они вычисляются путем деления соответствующей частоты на общее число наблюдений в выборке.

В рассматриваемом примере классы статистического ряда созданы для каждого значения признака. Однако если анализируемый признак варьирует в больших пределах, то целесообразно применять так называемую вторичную группировку данных. В этом случае в один класс объединяют несколько смежных значений признака. Например, при анализе распределения кладок соснового пилильщика по числу яиц [1] в первый класс относили кладки с числом яиц 10–24, во второй – 25–39, в третий – 40–54 и т. д. (табл. 6).

Таблица 6. **Распределение кладок пилильщика по числу яиц**

Классы (число яиц в кладке)	Частоты (число кладок в каждом классе)	Частости (доля кладок в каждом классе)
10–24	8	0,14
25–39	19	0,34
40–54	16	0,28
55–69	4	0,07
70–84	2	0,04
85–99	1	0,02
100–114	6	0,11
Итого	56	1,00

В случае количественной непрерывной вариации при группировке намечают классовые интервалы, а затем определяют число вариант, которые попали в каждый из интервалов.

Для того чтобы наметить классовые интервалы для равноинтервального вариационного ряда, необходимо интервал между минимальной и максимальной вариантами выборочной совокупности разделить на несколько одинаковых частей.

Если сделать малые классовые интервалы и большое число классов, то вариационный ряд получается слишком растянутым и повышается трудоемкость выполнения расчетов. Такой ряд распределения не дает четкой картины варьирования признака. Если, напротив, сделать слишком большие интервалы и малое число классов, то можно исказить характер варьирования. Кроме того, в этом случае снижается точность оценок статистических показателей, получаемых по материалам такого интервального вариационного ряда.

Следовательно, следует выбирать такое число классов и такую величину интервала, чтобы статистический ряд был хорошо обозримым, и в то же время обеспечивалась достаточная для целей проводимого исследования точность.

В литературе существует множество различных рекомендаций относительно выбора числа классов для интервальных статистических рядов [6, 18, 8, 16, 23, 7, 22, 19, 5, 1, 12, 14]. Проще всего для этой цели воспользоваться табл. 7 либо рассчитать число классов по формуле Стерджеса: $K = 1 + 3{,}322 \cdot \lg(n)$, а в случае, если n>100, можно воспользоваться формулой $K = 5 \cdot \lg(n)$ (Брукс С., Карузерс Н., 1963).

Таблица 7. **Рекомендуемое число классов**

Число наблюдений	Число классов
25–40	5–6
40–60	6–8
60–100	7–10
100–200	8–12
>200	10–15

После того как число классов будущего статистического ряда определено, можно рассчитать величину классового интервала:

$$\lambda = \frac{x_{\max} - x_{\min}}{K}, \tag{4}$$

где λ - величина классового интервала, $x_{\max}$ и $x_{\min}$ - максимальная и минимальная варианты; K – число классов.

Величину классового интервала округляют так, чтобы ее точность соответствовала точности представления анализируемых данных.

Величина классового интервала и число классов не определяют полностью границ интервалов вариационного ряда. Можно построить множество вариантов границ интервалов с одинаковой величиной классового интервала и числом классов, передвигая границы вдоль числовой оси.

Существуют различные предложения по выбору варианта размещения границ на числовой оси. Например, рекомендуется размещать границы таким образом, чтобы минимальное значение варианты в выборке попадало в середину первого интервала [12, 14]. В этом случае границы первого интервала можно определить следующим образом:

$$x_н^1 = x_{\min} - \lambda/2,$$
$$x_в^1 = x_{\min} + \lambda/2. \quad (5)$$

Границы остальных интервалов можно вычислить, пользуясь следующими соотношениями:

$$x_н^{i+1} = x_н^i + \lambda,$$
$$x_в^{i+1} = x_в^i + \lambda. \quad (6)$$

Существуют и другие рекомендации по поводу размещения на числовой оси интервалов вариационного ряда. Например, предлагается создавать такие границы интервалов, чтобы среднеарифметическое значение попадало в середину интервала [22]. Тогда границы одного из интервалов можно определить следующим образом:

$$x_н^k = \bar{x} - \lambda/2,$$
$$x_в^k = \bar{x} + \lambda/2, \quad (7)$$

где $x_н^k$ и $x_в^k$ - нижняя и верхняя границы k-го интервала (интервала, где находится среднеарифметическая величина), соответственно; $\bar{x}$ – приближенное значение среднеарифметического значения, оцененное по небольшой выборке. Границы остальных классовых интервалов можно определить на основании соотношений (6) и нижеследующих формул:

$$x_н^{i-1} = x_н^i - \lambda,$$
$$x_в^{i-1} = x_в^i - \lambda. \quad (8)$$

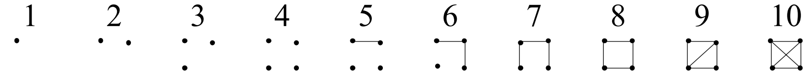
После того как границы классовых интервалов намечены, остается только разнести данные по классам.

Разноску частот по классам как интервального, так и безинтервального статистического ряда удобно выполнять следующим образом. Нужно подготовить таблицу, содержащую в первой колонке классы (границы классовых интервалов) (табл. 8). Во второй графе этой таблицы, просматривая результаты всех наблюдений, надо делать определенные пометки, регистрирующие попадание очередной варианты в соответствующий ей класс. Пометки удобно делать следующим образом. При попадании в класс первых четырех наблюдений эти события отмечают точками, как показано в табл. 9. Следующие шесть наблюдений фиксируют, соединяя всевозможные пары проставленных ранее четырех точек прямыми линиями. В результате после регистрации первых десяти наблюдений, попадающих в какой-нибудь интервал, во второй графе табл. 8 для соответствующего интервала появится изображение квадрата с двумя диагоналями («конверта»). Следующие десять наблюдений регистрируют таким же образом, как и первый десяток, и т. д. В конечном итоге после просмотра всей выборочной совокупности исходных данных во второй колонке табл. 8 в каждом классе будет присутствовать несколько полных и, как правило, один неполный «конверт». Каждый полный конверт дает десять наблюдений. Подсчитывая количество полных «конвертов», определяют число десятков, а подсчитывая число точек и черточек в неполном «конверте» – число единиц для частоты данного интервала. Полученные частоты интервалов заносят в третью колонку табл. 8.

Таблица 8. **Разноска вариант по классам**

Классы	Шифр частот	Частоты

Таблица 9. **Шифры частот, используемые при регистрации наблюдений**

Число наблюдений	1	2	3	4	5	6	7	8	9	10
Шифр частот										

В качестве примера составим интервальный вариационный ряд для выборочной совокупности диаметров, измеренных на высоте груди у 218 деревьев в чистом сосновом древостое (табл. 10).

Пользуясь табл. 7, выберем число интервалов для нашего вариационного ряда $K = 12$. Минимальная и максимальная варианта для данной выборочной совокупности диаметров следующие:

$x_{\min} = 9{,}8;$

$x_{\max} = 42{,}6.$

Воспользовавшись (4), определим величину классового интервала:

$$\lambda = \frac{x_{\max} - x_{\min}}{K} = \frac{42{,}6 - 9{,}8}{12} = 2{,}733 \approx 2{,}7. \qquad (9)$$

Далее, согласно (5), определим границы первого интервала:

$x_{н}^{1} = x_{\min} - \lambda / 2 = 9{,}8 - 2{,}7 / 2 = 8{,}45;$

$x_{в}^{1} = x_{\min} + \lambda / 2 = 9{,}8 + 2{,}7 / 2 = 11{,}15.$

Таблица 10. **Замеры высот и диаметров в чистом сосновом древостое**

Диаметр	Высота	Диаметр	Высота	Диаметр	Высота	Диаметр	Высота	Диаметр	Высота
23,2	20,5	31,6	24,5	29,7	14,5	19,9	19,5	29,7	24,7
30,7	23,0	34,2	24,0	14,7	24,5	29,2	22,5	26,9	25,7
28,3	23,5	14,4	14,5	29,0	23,5	20,8	21,5	32,6	24,7
26,9	25,5	25,5	17,5	20,8	14,5	33,9	24,5	21,4	23,1
25,4	21,5	25,0	20,5	32,0	24,5	14,4	16,5	33,4	25,5
30,8	22,5	28,8	22,5	31,8	23,5	23,7	21,5	31,0	25,7
20,1	18,5	23,9	21,5	21,1	21,5	20,9	17,0	23,2	23,9
29,1	24,5	27,3	21,5	18,6	18,5	26,6	21,5	25,0	26,8
29,3	23,5	15,8	17,5	20,6	18,5	22,5	20,5	17,7	21,7
15,9	18,5	21,1	22,5	28,4	20,5	17,9	21,7	34,5	24,7
23,1	21,0	18,0	20,5	22,9	21,5	33,5	26,2	31,4	25,7
24,4	21,0	32,4	23,0	25,8	23,5	27,3	23,7	26,4	25,2
25,3	21,0	13,8	16,5	24,0	22,0	23,0	24,7	27,0	22,7
24,8	22,0	30,8	23,0	21,5	23,5	31,4	22,7	24,8	23,7
34,3	23,0	17,8	18,5	12,5	16,5	24,9	22,7	28,2	24,7
27,6	19,5	26,3	20,5	18,9	24,5	28,0	22,2	29,2	21,2
24,1	20,5	10,3	11,5	12,5	16,5	26,6	23,7	24,6	24,9
23,7	20,5	31,1	22,5	20,2	21,5	27,6	24,7	14,6	17,7
35,8	22,0	10,7	11,0	28,1	23,5	29,6	24,2	23,6	23,9
20,1	21,5	28,3	19,5	19,2	21,5	20,8	22,7	28,7	24,9
21,6	21,5	17,5	17,5	24,3	20,5	24,3	21,7	31,5	25,7
17,8	21,5	13,5	16,5	24,0	24,5	19,5	19,7	20,6	18,8
19,1	22,5	19,9	18,0	19,6	19,5	25,6	22,7	19,3	18,7
23,4	24,5	23,3	21,5	19,6	22,5	34,9	26,7	24,4	23,9
28,5	25,5	21,5	22,0	24,8	20,0	19,3	19,7	28,4	21,7
14,4	17,5	22,0	22,0	25,0	23,5	39,5	21,7	25,5	23,2
26,9	19,5	21,4	19,5	15,9	14,5	19,9	21,7	25,4	22,2
14,1	17,5	18,8	22,5	31,5	20,5	25,3	20,7	23,2	22,7
23,6	21,5	29,3	21,5	22,9	19,5	16,5	15,7	17,7	20,7
37,5	26,0	24,8	22,5	42,6	25,5	25,2	21,7	25,0	21,5
31,4	21,5	30,8	19,5	9,8	17,5	18,5	13,7	18,1	17,0
32,9	23,5	19,1	20,5	14,9	18,5	26,5	21,7	27,2	20,7
31,2	22,5	30,0	23,7	22,6	21,5	30,7	22,7	22,2	21,7
34,5	24,5	23,3	20,5	28,0	23,5	20,7	19,2	23,8	22,7
28,7	24,5	27,0	20,0	24,1	21,5	13,6	13,7	22,4	22,7
24,9	23,0	17,7	18,5	15,6	16,5	22,4	24,2	24,0	25,2
26,4	22,5	15,7	19,5	32,1	24,5	21,1	23,7	22,6	25,7
16,9	19,5	28,5	16,5	28,3	18,5	22,6	22,7	25,8	26,9
18,9	22,0	20,5	21,0	28,9	24,5	33,6	24,7	28,0	24,2
23,9	21,5	18,7	22,0	25,0	23,5	23,3	21,7	25,5	22,7

Диаметр	Высота	Диаметр	Высота	Диаметр	Высота	Диаметр	Высота	Диаметр	Высота
25,4	22,5	21,2	18,5	23,2	26,5	27,1	21,7	22,3	20,2
19,0	18,5	14,7	21,5	13,9	20,5	26,6	21,9	18,9	21,9
11,3	15,5	20,7	15,5	16,3	17,5	33,5	25,8		
22,0	17,5	24,1	22,5	19,1	19,5	17,5	16,9		

Теперь, используя (6), можно определить границы всех классовых интервалов (табл. 11). Следующий этап составления вариационного ряда состоит в разноске наблюдений по классовым интервалам. Шифры частот, полученные в результате выполнения этой работы, приведены во второй колонке табл. 11. Третья и четвертая колонка данной таблицы содержит абсолютные и относительные частоты вариационного ряда по диаметрам.

Таблица 11. **Распределение деревьев по диаметру на высоте груди**

Классы	Шифр частот	Частоты	Относительные частоты	Накопленные частоты
8,45–11,15		3	0,0138	3
11,15–13,85		6	0,0275	9
13,85–16,55		16	0,0734	25
16,55–19,25		23	0,1055	48
19,25–21,95		29	0,1330	77
21,95–24,65		40	0,1834	117
24,65–27,35		39	0,1789	156
27,35–30,05		28	0,1284	184
30,05–32,75		19	0,0872	203
32,75–35,45		11	0,0505	214
35,42–38,15		2	0,0092	216
38,15–40,85		1	0,0046	217
40,85–43,55		1	0,0046	218
Итого		218	1	

В приведенном примере величина интервала 2,7 (9) оканчивается на нечетную цифру. Это привело к тому, что последняя цифра у всех границ интервалов – пять сотых. В связи с тем, что исходные данные представлены с точностью до десятых, при разноске наблюдений по интервалам всегда понятно, к какому интервалу следует отнести то или иное наблюдение. Попробуем построить интервалы для выборочной совокупности, представляющей собой измерения диаметров у 200 деревьев, взятых из последних четырех столбцов табл. 1 приложения с номерами 17 ,18, 19 и 20, в каждом из которых приведены замеры 50 деревьев. В рассматриваемой совокупности диаметров найдем минимальное и максимальное значения признака: $x_{\min} = 17{,}7$; $x_{\max} = 56{,}6$. С учетом объема выборки принимаем количество интервалов $k = 12$ в будущем статистическом ряду согласно рекомендациям, приведенным в табл. 7. Теперь, пользуясь формулой (4), определяем величину интервалов будущего вариационного ряда следующим образом:

$$\lambda = \frac{x_{\max} - x_{\min}}{k} = \frac{56{,}6 - 17{,}7}{12} = 3{,}2417 \text{ см} \approx 3{,}2 \text{ см}. \quad (10)$$

Далее определяем границы первого интервала вариационного ряда таким образом, чтобы минимальное значение в выборке попало в его середину (5):

$$x_{\text{н}}^1 = x_{\min} - \frac{\lambda}{2} = 17{,}7 - \frac{3{,}2}{2} = 16{,}1;$$

$$x_{\text{в}}^1 = x_{\min} + \frac{\lambda}{2} = 17{,}7 + \frac{3{,}2}{2} = 19{,}3;$$

Вычислив границы остальных интервалов, пользуясь формулами (6), получим: 16,1–19,3; 19,3–22,5; 22,5–25,7; 25,7–28,9; 28,9–32,1; 32,1–35,3; 35,3–38,5; 38,5–41,7; 41,7–44,9; 44,9–48,1; 48,1–51,3; 51,3–54,5; 54,5–57,6. Если делать разноску диаметров по приведенным выше интервалам, то в некоторых случаях будет возникать неопределенность (например, при разноске диаметра 25,7, так как он попадает точно на границу между третьим и четвертым интервалами). Для того чтобы избежать таких ситуаций, часто поступают следующим образом. Верхние границы интервалов уменьшают на величину, равную точности представления исходных данных. В этом случае нижние и верхние границы интервалов будут связаны следующем соотношением:

$$x_{\text{в}}^i = x_{\text{н}}^i + \lambda - \varepsilon, \quad (11)$$

где $x_{\text{в}}^i$ и $x_{\text{н}}^i$ - верхняя и нижняя границы i-того интервала; ε – минимальная величина, на которую могут отличаться различные значения признака (обусловливается точностью измерения данных).

Для того чтобы верхние и нижние границы были связаны соотношением (11), выражения (5) и (7) следует представить в следующем виде:

$$
\begin{aligned}
x_{н}^{1} &= x_{\min} - \lambda / 2; \\
x_{в}^{1} &= x_{\min} + \lambda / 2 - \varepsilon;
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
x_{н}^{k} &= \bar{x} - \lambda / 2; \\
x_{в}^{k} &= \bar{x} + \lambda / 2 - \varepsilon.
\end{aligned}
\tag{13}
$$

Так как измерения диаметров в рассматриваемом примере выполнены с точностью до 0,1 см, то значение ε следует взять равным $\varepsilon = 0{,}1$. Новые, сформированные с помощью выражений (12) и (6) интервалы и результаты разноски по ним диаметров деревьев, приведены в табл. 12.

В качестве значений для сформированных классовых интервалов в дальнейших расчетах используют их середины, которые можно вычислить по следующей формуле:

$$
x_i = \frac{x_{н}^{i} + x_{в}^{i}}{2}, \tag{14}
$$

где x_i – значения классовых интервалов.

Вычисленные по формуле (14) середины интервалов для рассматриваемого примера приведены в третьей колонке табл. 12.

Кроме описанной выше простой группировки данных, в некоторых случаях группировку целесообразно осуществить сразу по двум параметрам. Такая необходимость, как правило, возникает, если нужно проанализировать связь между двумя случайными величинами. В таких случаях составляют так называемые *таблицы распределения* (корреляционные таблицы, корреляционные решетки).

Для примера составим таблицу распределения деревьев по классам диаметра и высоты в чистом сосновом древостое по материалам выборочной совокупности, представляющей собой измерения диаметров и высот у 200 деревьев, взятых из последних четырех столбцов табл. 1 приложения с номерами 17 ,18, 19 и 20. Для этого воспользуемся интервалами рядов распределения деревьев по диаметрам (табл. 12).

Так же как и для ряда по диаметрам, пользуясь табл. 7, выберем число интервалов K для вариационного ряда по высотам, равным 12. Минимальная и максимальная высота в рассматриваемом примере равны: $x_{\min} = 17{,}5$; $x_{\max} = 29{,}6$.

Далее, согласно (4), находим величину классового интервала:

$$\lambda = \frac{x_{\max} - x_{\min}}{k} = \frac{29{,}6 - 17{,}5}{12} = 1{,}0083\,\text{см} \approx 1{,}0\,\text{см}, \tag{15}$$

Таблица 12. **Распределение наблюдений по интервалам (диаметры)**

Классы	Шифр частот	Средние значения классов	Частоты	Накопленные частоты
16,1–19,2		17,65	3	3
19,3–22,4		20,85	11	14
22,5–25,6		24,05	29	43
25,7–28,8		27,25	39	82
28,9–32,0		30,45	32	114
32,1–35,2		33,65	33	147
35,3–38,4		36,85	23	170
38,5–41,6		40,05	10	180
41,7–44,8		43,25	9	189
44,9–48,0		46,45	3	192
48,1–51,2		49,65	4	196
51,3–54,4		52,85	2	198
54,5–57,6		56,05	2	200
Итого			200	

Построим интервалы для ряда высот таким образом, чтобы среднее арифметическое значение попало в середину интервала. Для этого воспользуемся выражениями (13). В связи с тем, что высоты в данном примере измерены с точностью до 0,1 м, значение ε установим равным $\varepsilon = 0{,}1$. Оценку среднего арифметического вычислим, просуммировав первые 25 высот из рассматриваемой выборки и разделив их на 25:

$$\bar{x} = \frac{651{,}8}{25} = 26{,}072 \approx 26{,}1.$$

Далее, воспользовавшись формулами (13), получим:

$$x_{\text{н}}^{k} = \bar{x} - \lambda / 2 = 26{,}1 - 1{,}0 / 2 = 25{,}6;$$

$$x_{\text{в}}^{k} = \bar{x} + \lambda / 2 - \varepsilon = 26{,}1 + 1{,}0 / 2 - 0{,}1 = 26{,}5.$$

Используя (6) и (8), можно определить границы всех классовых интервалов. Полученные интервалы и частоты вариационного ряда высот приведены в табл. 13.

Таблица 13. **Распределение наблюдений по интервалам (высоты)**

Классы	Шифр частот	Средние значения классов	Частоты	Накопленные частоты
16,6–17,5		17,05	2	2
17,6–18,5		18,05	1	3
18,6–19,5		19,05	4	7
19,6–20,5		20,05	5	12
20,6–21,5		21,05	2	14
21,6–22,5		22,05	13	27
22,6–23,5		23,05	25	52
23,6–24,5		24,05	31	83
24,6–25,5		25,05	32	115
25,6–26,5		26,05	43	158
26,6–27,5		27,05	24	182
27,6–28,5		28,05	14	196
28,6–29,5		29,05	3	199
29,6–30,5		30,05	1	200
Итого			200	

Для построения корреляционной решетки вычисленные границы интервалов по высоте заносятся в первую колонку, а границы интервалов ряда по диаметрам – в первую строку таблицы (корреляционной решетки) (табл. 14). Далее необходимо выполнить разноску деревьев по клеткам корреляционной таблицы на основании значений их высот и диаметров. Шифры частот, полученные в результате выполнения этой работы, заносятся в клетки таблицы, а затем под дробной чертой записываются полученные численности для каждой клетки корреляционной решетки.

Табл. 14. **Распределение наблюдений по интервалам диаметра и высоты**

H \ D	16,1–19,2	19,3–22,4	22,5–25,6	25,7–28,8	28,9–32,0	32,1–35,2	35,3–38,4	38,5–41,6	41,7–44,8	44,9–48,0	48,1–51,2	51,3–54,4	54,5–57,6	Всего
29,6–30,5				/1										1
28,6–29,5				/1	/1						/1			3
27,6–28,5			/1	/1	/2	/1	/3	/2		/1	/2	/1		14
26,6–27,5				/1	/3	/4	/5	/3	/4	/2			/2	24
25,6–26,5			/1	/7	/4	/13	/10	/3	/4		/1			43
24,6–25,5			/1	/7	/7	/12	/2	/1	/1			/1		32
23,6–24,5			/7	/7	/12	/1	/3	/1						31
22,6–23,5		/1	/8	/12	/2	/2								25
21,6–22,5		/3	/8	/1	/1									13
20,6–21,5		/1		/1										2
19,6–20,5		/4	/1											5
18,6–19,5	/1	/2	/1											4
17,6–18,5			/1											1
16,6–17,5	/2													2
Итого	3	11	29	39	32	33	23	10	9	3	4	2	2	200

Проанализировав полученную таблицу распределения деревьев в чистом одновозрастном сосновом древостое по высотам и диаметрам, можно прийти к выводу, что данные в таблице расположились не хаотично, а подчиняясь некоторым закономерностям. Хорошо видно, что в табл. 14 остались пустыми левый верхний угол и правый нижний. И не удивительно, ведь левый верхний угол соответствует деревьям, которые имеют очень большую высоту и очень маленький диаметр. Таких очень тонких и высоких деревьев в лесу нет. Правый нижний угол соответствует очень толстым и очень низким деревьям. Но в обычном сосновом древостое нет деревьев, имеющих такую форму ствола.

Основная масса наблюдений расположилась в табл. 14 по диагонали от левого нижнего угла к верхнему правому. Такой характер расположения данных в корреляционной решетке говорит нам о том, что, по-видимому, между высотами и диаметрами деревьев в древостое имеется связь, а именно, при увеличении диаметра деревьев увеличивается и их высота. Кроме того, так как данные расположились в таблице не ровно на диагонали, а с некоторым отклонением от нее, можно предположить, что эта связь носит нелинейный характер.

Однако такой анализ корреляционной решетки слишком субъективен и не может дать исчерпывающей информации о связи между анализируемыми признаками. Для более детального изучения зависимостей между параметрами целесообразно изобразить таблицу распределения графически, а также воспользоваться возможностями, предоставляемыми корреляционным и регрессионными анализами. Более подробно эти вопросы будут рассмотрены в последующих главах.

1.5. Эмпирическая функция распределения

Эмпирическая функция распределения является очень важной характеристикой экспериментальных данных. Она задается следующими соотношениями:

$$F_n(x) = \begin{cases} 0, x < x_1^{(n)}, \\ k/n, x_k^{(n)} \le x < x_{k+1}^{(n)}, \\ 1, x \ge x_n^{(n)}, \end{cases} \qquad (16)$$

30

или

$$F_n(x) = \frac{1}{n} \cdot \sum_{k=1}^{n} e(x - x_k^{(n)}),$$

где $e(x)$ – функция единичного скачка (функция Хевисайда):

$$e(x) = \begin{cases} 0, x < 0, \\ 1, x \geq 0. \end{cases}$$

На рис. 1 приведен типичный график функции $F_n(x)$.

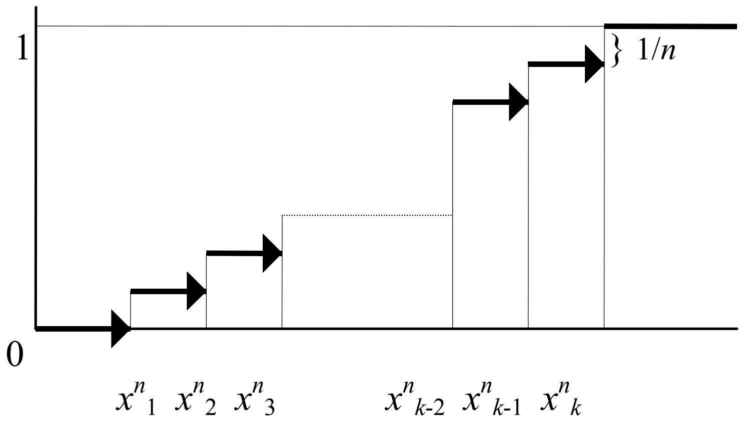


Рис. 1. Эмпирическая функция распределения

Очень важным является то, что эмпирическая функция распределения $F_n(x)$ при увеличении размера выборки n ($n \to \infty$) сходится по вероятности к функции распределения $F(x)$ генеральной совокупности. В связи с этим $F_n(x)$ используется для оценки функции распределения $F(x)$.

Эмпирическая функция распределения $F_n(x)$ обладает следующими свойствами:

1. Значения функции $F_n(x)$ принадлежат отрезку [0; 1].

2. Эмпирическая функция распределения является неубывающей функцией.

3. Функция $F_n(x)$ непрерывна справа.

4. $F_n(x) = 0$, если $x < x_1^n$ и $F_n(x) = 1$, если $x \geq x_n^n$, где x_1^n – минимальное значение в выборке, а x_n^n – максимальное.

31

Определим значения эмпирической функции распределения $F_n(x)$ на основании данных интервальных вариационных рядов по диаметрам и высотам деревьев в сосновом древостое (табл. 12 и 13). Для выполнения вычислений нам понадобятся накопленные частоты рядов распределения, которые, согласно (16), необходимо разделить на число наблюдений в выборке. Результаты вычисления значений эмпирической функции распределения приведены в табл. 15.

1.6. Графическое изображение статистических рядов

Графическое изображение вариационных рядов позволяет увидеть и проанализировать особенности варьирования изучаемого признака. Для этой цели используются различные типы графиков. При анализе массовых наблюдений для наглядного представления плотности распределения изучаемой случайной величины чаще всего применяются гистограмма и полигон распределения. Увидеть эмпирическую функцию распределения позволяет кумулята. Кроме кумуляты, для графического представления вариационных рядов используется огива. Для изображения таблиц распределения (корреляционных решеток) используют трехмерные графики, называемые *призмограммами*. Рассмотрим эти виды графиков.

Таблица 15. **Эмпирические функции распределения диаметров и высот деревьев в сосновом древостое**

Диаметры				Высоты			
x_i	Частоты	Накоп-ленные частоты	Эмпирическая функция распределения	x_i	Частоты	Накопленные частоты	Эмпирическая функция распределения
17,65	3	3	0,015	17,05	2	2	0,010
20,85	11	14	0,070	18,05	1	3	0,015
24,05	29	43	0,215	19,05	4	7	0,035
27,25	39	82	0,410	20,05	5	12	0,060
30,45	32	114	0,570	21,05	2	14	0,070
33,65	33	147	0,735	22,05	13	27	0,135
36,85	23	170	0,850	23,05	25	52	0,260
40,05	10	180	0,900	24,05	31	83	0,415
43,25	9	189	0,945	25,05	32	115	0,575
46,45	3	192	0,960	26,05	43	158	0,790

32

Диаметры				Высоты			
x_i	Частоты	Накоп-ленные частоты	Эмпирическая функция распределения	x_i	Частоты	Накопленные частоты	Эмпирическая функция распределения
49,65	4	196	0,980	27,05	24	182	0,910
52,85	2	198	0,990	28,05	14	196	0,980
56,05	2	200	1,000	29,05	3	199	0,995
				30,05	1	200	1,000
$\sum$	200			$\sum$	200		

Гистограмма. Гистограммы распределения строятся только для интервальных вариационных рядов. При этом на оси абсцисс откладывают границы интервалов. График изображается в виде прямоугольников, основаниями которых служат интервалы ряда. Высоты прямоугольников должны соответствовать частотам (или частостям), откладываемым по оси ординат. На рис. 2 приведен пример гистограммы распределения диаметров в сосновом древостое, построенный по данным вариационного ряда, приведенного в табл. 11.

Полигон. Полигон распределения может строиться как для интервальных вариационных рядов, так и для дискретных. При его построении на горизонтальной оси откладывают значения классов, а по вертикальной – частоты или частости (рис. 3). Затем полученные точки соединяют прямыми линиями. При построении полигона используют два дополнительных примыкающих к вариационному ряду интервала, имеющих нулевые частоты. На рис. 3 изображен полигон, построенный для распределения деревьев в древостое по диаметрам для вариационного ряда, приведенного в табл. 11.

33

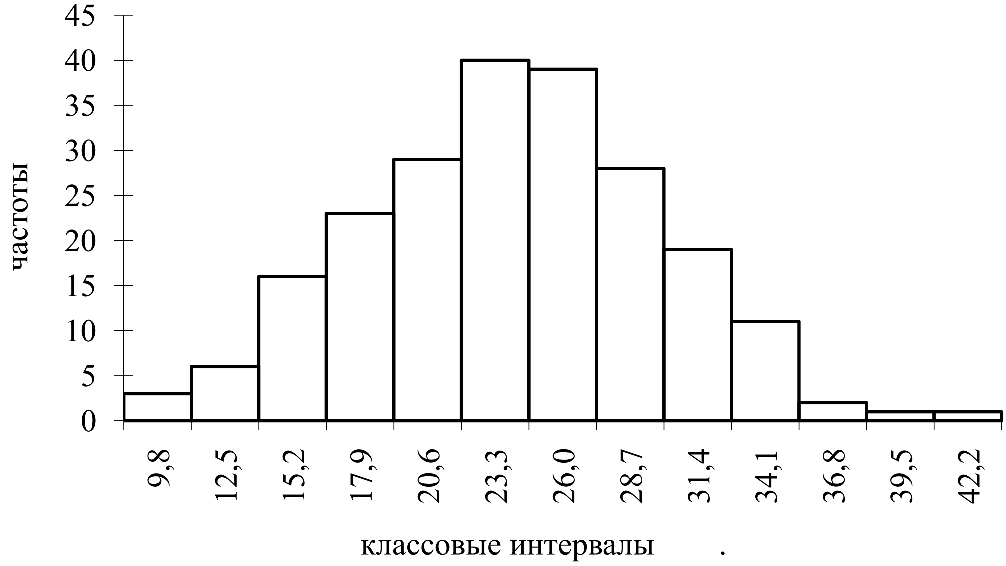


Рис. 2. Гистограмма распределения 218 сосновых стволов по диаметрам

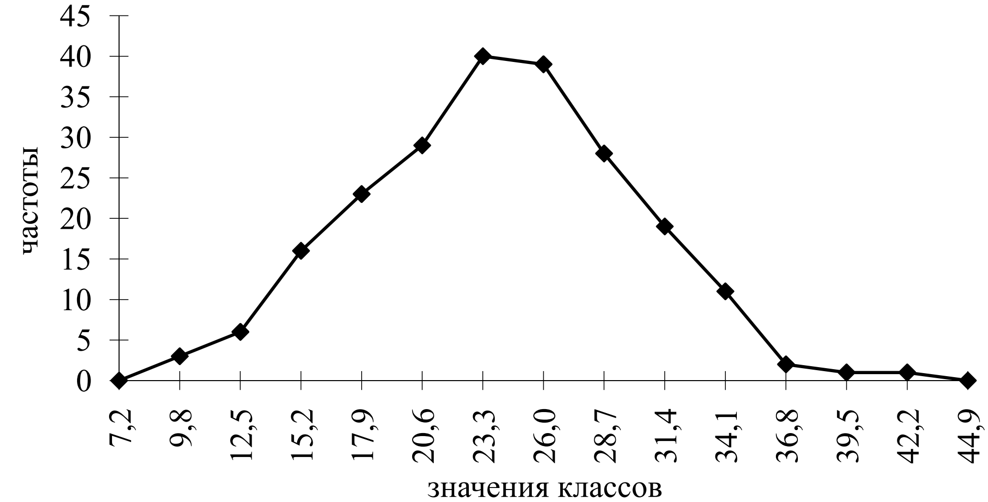


Рис. 3. Полигон распределения 218 сосновых стволов по диаметрам

Кумулята. При построении кумулятивной кривой по оси ординат откладывают накопленные частоты или частости. По оси абсцисс для дискретного вариационного ряда откладывают значения классов (рис. 4), а для непрерывного – границы интервалов (рис. 5). В

34

последнем случае накопленные частоты или частости соответствуют верхним границам интервалов. Нижней границе первого интервала соответствует нулевое значение накопленной частоты или частости. Полученные точки затем соединяют прямыми линиями.

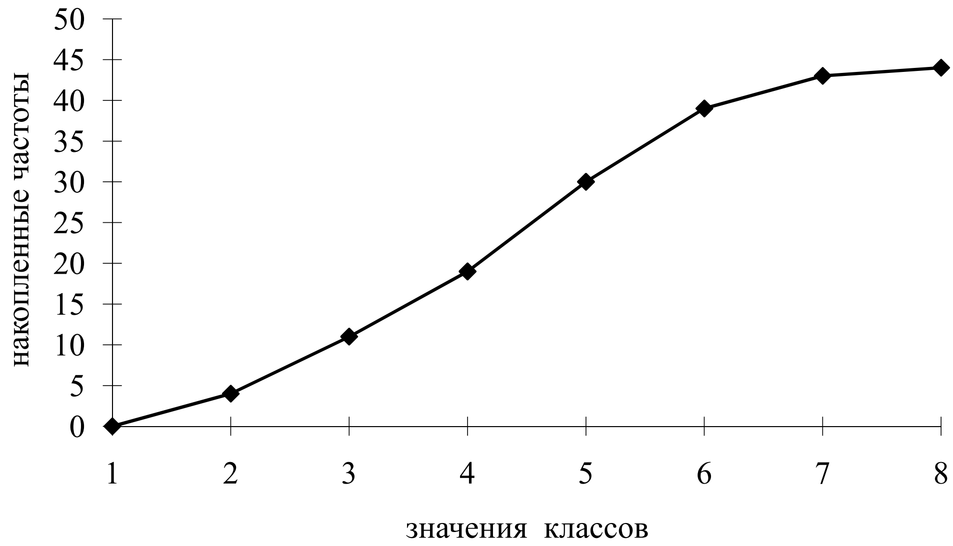


Рис. 4. Кумулята распределения сеголеток (первые две генерации) рыжей полевки по величине сентябрьских выводков

35

Рис. 5. Кумулята распределения 218 сосновых стволов по диаметрам

Огива. Огива строится так же, как и кумулята, однако координаты меняются ролями. На оси абсцисс откладываются частоты или частости, а значения классов или границы интервалов – по оси ординат. Пример огивы, построенной по данным ряда диаметров, приведенного в табл. 11, показан на рис. 6.

Наиболее выразительными графики получаются, если при их построении соблюдать правило золотого сечения, согласно которому основание графика должно относиться к его высоте как 8:5 = 1:0,625.

Гистограмма и полигон иллюстрируют плотность распределения вероятностей изучаемой случайной величины. Они имеют куполообразную форму. Их удобно использовать при подборе теоретического распределения для исследуемых данных.

Кумулята и огива, в отличие от первых двух типов графиков, обычно имеют более плавные контуры. Кумулята предоставляет возможность увидеть эмпирическую функцию распределения анализируемой случайной величины.

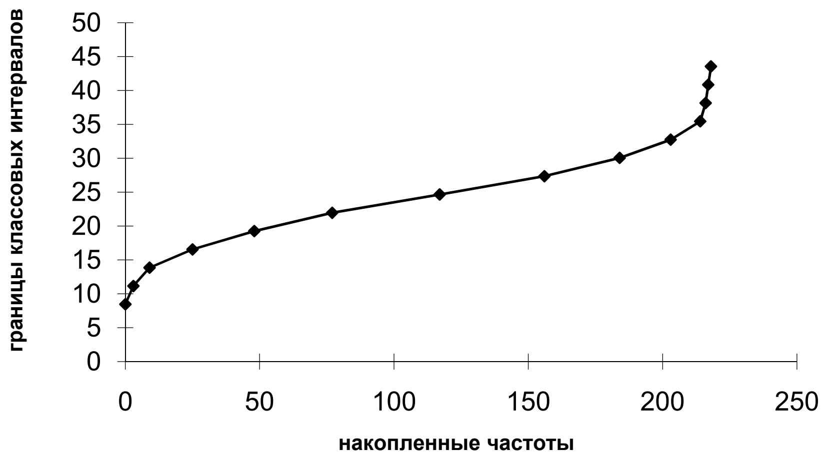


Рис. 6. Огива распределения 218 сосен по диаметрам на высоте груди

Призмограмма. В том случае, если исследуется связь между двумя случайными величинами, т. е. анализируются сразу два признака и составляется корреляционная решетка, изобразить таблицу распределения можно в виде призмограммы. Этот тип графика представляет собой трехмерный аналог гистограммы. На рис. 7 приведен пример призмограммы, построенной по данным распределения деревьев в чистом одновозрастном сосновом древостое по диаметрам и высотам (табл. 14).

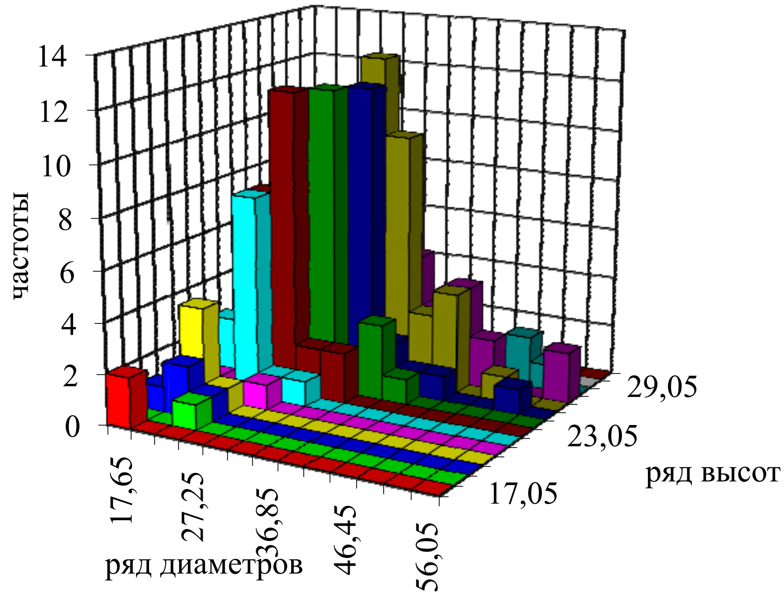


Рис. 7. Призмограмма распределения деревьев по классам диаметра и высоты в чистом сосновом древостое

При построении призмограммы по оси *X* и *Y* откладывают значения признаков (в рассматриваемом примере это диаметры и высоты). По оси *Z* откладывают частоты корреляционной решетки.

37

2. ОСНОВНЫЕ ПОКАЗАТЕЛИ ВЫБОРОЧНОЙ СОВОКУПНОСТИ

Группировка исходных данных и представление их в виде вариационных рядов позволяет представить статистические данные в компактном и обозримом виде. Представление статистических рядов в графическом виде дает информацию о характере распределения изучаемой случайной величины. Однако наряду с качественной характеристикой статистических данных часто необходимы количественные показатели, обобщающие информацию, содержащуюся в исходных данных.

Такие показатели можно разделить на несколько групп в зависимости от того, каким образом они характеризуют экспериментальные данные.

Первая группа показателей указывает на местоположение значений изучаемого признака на числовой оси. Они определяют ту величину, вокруг которой сосредоточены все значения изучаемого признака. К этой группе показателей относятся всевозможные средние величины.

Вторая группа характеристик выборки показывает, каким образом значения анализируемого признака располагаются вокруг средних значений. Эти показатели характеризуют степень изменчивости исходных данных. Они называются показателями вариации.

Третья группа объединяет в себе показатели, называющиеся структурными характеристиками вариационного ряда. Статистические показатели из этой группы характеризуют как расположение значений изучаемого признака на числовой оси, так и характер распределения изучаемой случайной величины.

2.1. Средние величины

Как уже отмечалось выше, средние величины указывают то значение, вокруг которого группируются величины изучаемого признака. Средние величины могут быть *степенными*, которые будут рассматриваться в данном разделе, и *структурными*, которые мы рассмотрим вместе с остальными структурными характеристиками вариационных рядов.

Все степенные средние могут быть заданы с помощью формулы

$$\bar{x} = \left(\frac{\sum x_i^m}{n} \right)^{1/m} = \sqrt[m]{\frac{\sum x_i^m}{n}},$$

где $\bar{x}$ – средняя величина; x_i – варианта; n – число наблюдений, для которых вычисляют среднюю.

Средние величины характеризуют различные свойства изучаемой совокупности данных. Свойство, представляемое какой-либо из средних, называется *определяющим*. Это понятие впервые было введено А. Я. Боярским. Наиболее распространенной средней величиной является среднее арифметическое. Определяющее свойство для этой средней заключается в том, что если каждое наблюдение в выборке заменить средней арифметической, то сумма результатов наблюдений не изменится, т. е.

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \bar{x}, \tag{17}$$

где x_i – i-тое наблюдение; $\bar{x}$ – средняя арифметическая; n – объем выборки. Учитывая, что $\bar{x} = const$, (17) можно переписать в виде

$$\sum_{i=1}^{n} x_i = n \cdot \bar{x},$$

откуда

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{18}$$

Таким образом, мы получили формулу для вычисления средней арифметической величины.

В общем случае определяющее свойство для степенных средних заключается в том, что если каждое отдельное наблюдение заменить средней величиной, то сумма m-ных степеней наблюдений не изменится. Это можно записать в виде равенства:

$$\sum_{i=1}^{n} x_i^m = \sum_{i=1}^{n} \bar{x}_m^m, \tag{19}$$

где m – порядок степенной средней, $\bar{x}_m$ - степенная средняя m-ного порядка. Преобразовав выражение (19), можно получить формулу для определения степенной средней m-го порядка:

$$\bar{x}_m = \sqrt[m]{\frac{\sum_{i=1}^{n} x_i^m}{n}} = \left(\frac{\sum_{i=1}^{n} x_i^m}{n} \right)^{1/m}. \qquad (20)$$

Теперь рассмотрим некоторые наиболее употребительные степенные средние.

Средняя арифметическая. Эта величина является степенной средней 1-го порядка. В соответствии с определяющим свойством при замене всех вариант средним арифметическим значением сумма наблюдений не изменится.

Если исходные данные не сгруппированы, то среднее арифметическое можно вычислить согласно (18). В том случае, если этот показатель вычисляется на основании сгруппированного набора данных, следует воспользоваться формулой

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n}. \qquad (21)$$

где k – количество классов; x_i – значение i-того класса (середина интервала, если ряд интервальный); f_i – частота i-того класса. Средняя арифметическая величина, вычисленная по формуле (21), называется *взвешенной*.

Средняя арифметическая величина обладает рядом свойств.

1. Сумма отклонений вариант от средней арифметической величины равна нулю. Действительно,

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n \cdot \bar{x}$$

Используя (18), получим

$$\sum_{i=1}^{n} x_i - n \cdot \bar{x} = \sum_{i=1}^{n} x_i - n \cdot \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0.$$

2. Если к каждому наблюдению прибавить константу c, то и среднее арифметическое изменится соответствующим образом. Покажем это. Прибавляя к левой и правой частям равенства (18) константу c и преобразуя правую часть полученного равенства, имеем

$$\overline{x}+c=\frac{\sum_{i=1}^{n}x_i}{n}+c=\frac{\sum_{i=1}^{n}x_i+n\cdot c}{n}=\frac{\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}c}{n}=\frac{\sum_{i=1}^{n}(x_i+c)}{n}=\overline{x+c}. \quad (22)$$

3. Если каждое наблюдение умножить на константу c, то среднее арифметическое изменится аналогичным образом. Действительно, если умножить левую и правую части равенства (18) на c и преобразовать правую части полученного равенства, имеем

$$c\cdot\overline{x}=c\cdot\frac{\sum_{i=1}^{n}x_i}{n}=\frac{\sum_{i=1}^{n}c\cdot x_i}{n}.$$

4. Сумма квадратов отклонений вариант от их средней арифметической величины меньше, чем сумма квадратов отклонений от любой другой величины, не равной $\overline{x}$, т. е.

$$\sum(x_i-\overline{x})^2<\sum(x_i-a)^2 \text{ для любого } a\neq\overline{x}. \quad (23)$$

Чтобы доказать это, покажем, что выражение

$$\sum(x_i-a)^2 \quad (24)$$

достигает минимума в точке $a=\overline{x}$. Продифференцировав (24) по a, получим

$$\frac{d\left(\sum(x_i-a)^2\right)}{da}=2\cdot\sum(x_i-a)\cdot(-1).$$

Далее найдем, в какой точке производная от выражения (24) принимает нулевое значение:

$$2\cdot\sum(x_i-a)\cdot(-1)=0,$$

откуда

$$\sum x_i-n\cdot a=0$$

и

$$a=\frac{\sum x_i}{n}=\overline{x}.$$

Отсюда следует, что сумма квадратов отклонений (24) имеет один локальный экстремум в точке $a=\overline{x}$. Так как вторая производная этой функции по a в точке $a=\overline{x}$ равна положительному числу

$$\frac{d\left(2\cdot\sum(x_i-a)\cdot(-1)\right)}{da}=2,$$

то в точке $a = \bar{x}$ функция (24) имеет минимум.

5. Средняя арифметическая величина ряда, состоящего из двух групп наблюдений, равна средневзвешенной из средних значений групп наблюдений с весами, равными объемам групп:

$$\bar{x} = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2}{n_1 + n_2}. \tag{25}$$

Покажем это. Используя (18), правую часть (25) можно преобразовать следующим образом:

$$\frac{\frac{\sum_{i=1}^{n_1} x_i}{n_1} \cdot n_1 + \frac{\sum_{i=n_1+1}^{n_1+n_2} x_i}{n_2} \cdot n_2}{n_1 + n_2} = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=n_1+1}^{n_1+n_2} x_i}{n_1 + n_2} = \frac{\sum_{i=1}^{n_1+n_2} x_i}{n_1 + n_2} = \bar{x}.$$

В общем случае, когда ряд наблюдений состоит из k групп, средняя арифметическая величина всего ряда равна средневзвешенной арифметической групповых средних с весами, равными объемам групп:

$$\bar{x} = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2 + \ldots + \bar{x}_k \cdot n_k}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{i=1}^{k} \bar{x}_i \cdot n_i}{\sum_{i=1}^{k} n_i}.$$

6. Средняя арифметическая величина для суммы или разности взаимно соответствующих значений признаков двух рядов наблюдений с одинаковым числом наблюдений равна сумме или разности средних арифметических значений этих рядов. Действительно, используя (18), мы можем записать

$$\overline{x \pm y} = \frac{\sum_{i=1}^{n} (x_i \pm y_i)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \pm \frac{\sum_{i=1}^{n} y_i}{n} = \bar{x} \pm \bar{y}.$$

Распространяя данное свойство на сумму или разность k групп с одинаковым числом наблюдений, можно сказать, что их общая средняя арифметическая равна сумме или разности средних арифметических групп:

$$\overline{x^1 \pm x^2 \pm \ldots \pm x^k} = \frac{\sum_{i=1}^{n} (x_i^1 \pm x_i^2 \pm \ldots \pm x_i^k)}{n} =$$

$$= \frac{\sum_{i=1}^{n} x_i^1}{n} \pm \frac{\sum_{i=1}^{n} x_i^2}{n} \pm \ldots \pm \frac{\sum_{i=1}^{n} x_i^k}{n} = \overline{x^1} \pm \overline{x^2} \pm \ldots \pm \overline{x^k}.$$

В качестве примера определим среднее арифметическое значение прироста по высоте за десятилетние периоды роста дерева. Исходные данные приведены в табл. 16. Используя (18), получим

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{29\ \text{м}}{10} = 2{,}9\ \text{м}.$$

Таким образом, в среднем за десятилетний период высота дерева увеличивалась на 2,9 м. Зная среднее значение приростов за десятилетние периоды, нетрудно узнать высоту дерева в 100-летнем возрасте:

$$2{,}9 \cdot 10 = 29\ \text{м}.$$

Это значение в точности равно сумме всех значений прироста за десятилетние периоды времени (табл. 16), что соответствует определяющему свойству средней арифметической величины.

Средняя гармоническая. Это величина является степенной средней порядка $m = -1$. Подставляя это значение m в (20), получим

$$\bar{x}_{-1} = \left(\frac{\sum_{i=1}^{n} x_i^{-1}}{n} \right)^{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}. \qquad (26)$$

Определяющее свойство для этой степенной средней гласит: при замене всех значений средним геометрическим сумма обратных величин наблюдений не изменится.

Формула (26) пригодна для вычисления средней гармонической в том случае, если исходные данные не сгруппированы. В том случае, если этот показатель рассчитывается по данным статистического ряда, используют формулу

Таблица. 16. **Ход роста древесного ствола**

Возраст дерева, лет	Высота дерева, м	Прирост по высоте, м
0	0	
		0,9
10	0,9	
		1,7

20	2,6	
		4,0
30	6,6	
		4,6
40	11,2	
		2,4
50	13,6	
		3,4
60	17,0	
		3,7
70	20,7	
		3,1
80	23,8	
		3,2
90	27,0	
		2,0
100	29,0	

$$\bar{x}_{-1} = \frac{n}{\sum_{i=1}^{k} \left(f_i / x_i \right)}. \qquad (27)$$

Рассмотрим использование средней гармонической на примере. Предположим, что для обработки почвы при лесовосстановительных мероприятиях в лесхозе имеется две зубовые бороны ЗБЗТУ-1, по одной зубовой бороне ЗБЗС-1 и ЗБП-0,6 и три зубовые бороны ШБ-2,5. Анализ производительности этих орудий показал, что с помощью бороны ЗБЗТУ–1 или ЗБЗС–1 для обработки 1 га почвы необходимо 0,5 ч, с помощью бороны ЗБП-0,6 – 0,71 ч, с помощью бороны ШБ-2,5 – 0,67 ч. Найдем, сколько в среднем необходимо времени на обработку 1 га почвы с помощью орудий, имеющихся в лесхозе.

Если для этой цели использовать среднюю арифметическую, то получим следующий результат:

$$\frac{2 \cdot 0{,}50 + 0{,}50 + 0{,}71 + 3 \cdot 0{,}67}{7} = 0{,}603 \text{ ч.}$$

Теперь, пользуясь полученным средним значением, попытаемся определить, сколько лесхоз с помощью имеющихся в наличии орудий может обработать почвы за 6 ч:

$$\frac{6}{0{,}60}\cdot 7 = 69{,}7 \text{ га.}$$

В действительности за 6 ч с помощью имеющихся орудий в лесхозе можно обработать следующую площадь:

$$2\cdot\frac{6}{0{,}50}+\frac{6}{0{,}50}+\frac{6}{0{,}71}+3\cdot\frac{6}{0{,}67} = 71{,}3 \text{ га.}$$

Теперь попробуем оценить среднее время, необходимое для обработки 1 га почвы с помощью имеющихся в лесхозе орудий, используя среднюю гармоническую. Получим

$$\frac{7}{\frac{2}{0{,}50}+\frac{1}{0{,}50}+\frac{1}{0{,}71}+\frac{3}{0{,}67}} = 0{,}589 \text{ ч.}$$

Определим, сколько лесхоз с помощью имеющихся в наличии орудий может обработать почвы за 6 ч, пользуясь средней гармонической величиной:

$$\frac{6}{0{,}59}\cdot 7 = 71{,}3 \text{ га.}$$

Как видим, в данном случае для оценки среднего времени, необходимого для обработки почвы орудиями, имеющимися в лесхозе, предпочтение следует отдать средней гармонической величине.

Средняя квадратическая. Подставим в (20) значение $m = 2$:

$$\bar{x}_2 = \sqrt[2]{\frac{\sum_{i=1}^{n} x_i^2}{n}}.$$

Полученная средняя называется средней квадратической. Согласно определяющему свойству для этой величины при замене всех значений средним квадратическим сумма квадратов наблюдений не изменится.

Если среднюю квадратическую необходимо вычислить для сгруппированного набора данных, то следует применить формулу

$$\bar{x}_2 = \sqrt[2]{\frac{\sum_{i=1}^{k} f_i \cdot x_i^2}{n}}. \qquad (28)$$

В качестве примера вычислим среднюю квадратическую величину для ряда распределения стволов по диаметру на высоте груди (табл. 11). В табл. 17 приведены вспомогательные расчеты для разбираемого примера. Подставляя в (28) суммы из 2-й и 4-й колонок данной таблицы, получим

$$\bar{x}_2 = \sqrt{\frac{133314}{218}} = \sqrt{611{,}53} = 24{,}73\,\text{см}.$$

Таблица 17. **Вспомогательные расчеты для определения средней квадратической**

x_i	f_i	x_i^2	$f_i \cdot x_i^2$	g_i	$f_i \cdot x_i$
1	2	3	4	5	6
9,8	3	96,04	288	0,0226	29,4
12,5	6	156,25	938	0,0736	75,0
15,2	16	231,04	3697	0,2903	243,2
17,9	23	320,41	7369	0,5788	411,7
20,6	29	424,36	12 306	0,9665	597,4
23,3	40	542,89	21 716	1,7055	932,0
26,0	39	676,00	26 364	2,0706	1014,0
28,7	28	823,69	23 063	1,8114	803,6
31,4	19	985,96	18 733	1,4713	596,6
34,1	11	1162,81	1271	1,0046	375,1
36,8	2	1354,24	2708	0,2127	73,6
39,5	1	1560,25	1560	0,1225	39,5
42,2	1	1780,84	1781	0,1399	42,2
Итого	218		133 314	10,4703	5233,3

В лесном хозяйстве очень важное значение имеет такой показатель, как сумма площадей сечений стволов на высоте груди. Попробуем вычислить этот параметр, пользуясь средней квадратической величиной:

$$G = n \cdot \frac{\pi \cdot D^2}{4} = 218 \cdot \frac{3{,}14 \cdot 24{,}73}{4} = 104\,658\,\text{см}^2 = 10{,}4658\,\text{м}^2,$$

где G – сумма площадей сечений; D – среднеквадратический диаметр.

Для сравнения в 5-й колонке вспомогательной таблицы (24) рассчитаны действительная площадь сечения стволов на высоте груди анализируемого древостоя. В данной колонке приводятся площади сечений деревьев, попавших в каждый из интервалов вариационного ряда, вычисленные по формуле

$$g_i = f_i \cdot \frac{\pi \cdot x_i^2}{4} \cdot 0{,}0001.$$

Коэффициент 0,0001 в данной формуле служит для перехода от сантиметров, в которых измерялись все диаметры в данном примере, к квадратным метрам, в которых в лесном хозяйстве принято измерять сумму площадей сечений. Сумма чисел, занесенных в данную колонку, и даст нам сумму площадей сечений анализируемого древостоя.

Как видим, полученные разными способами площади сечений практически совпадают, отличаясь только в последних разрядах, что вызвано округлением чисел в ходе вычислений.

Попробуем вычислить сумму площадей сечений, пользуясь средней арифметической величиной. Подставляя суммы из 2-й и 6-й колонок вспомогательной табл. 17 в формулу (21), вычислим среднюю арифметическую величину:

$$\bar{x}_1 = \frac{5233{,}3}{218} = 24{,}01\,\text{см}.$$

Далее на основании средней арифметической определим сумму площадей поперечных сечений стволов древостоя:

$$G = n \cdot \frac{\pi \cdot D^2}{4} = 218 \cdot \frac{3{,}14 \cdot 24{,}01}{4} = 98\,653\,\text{см}^2 = 9{,}8653\,\text{м}^2.$$

Как видим, полученное значение отличается от истинного. Следовательно, если в проводимых статистических исследованиях важное значение имеет площадь сечения, то в таком случае наиболее целесообразно использовать среднюю квадратическую величину.

Средняя кубическая. В том случае, если учитывался признак, характеризующий линейный размер объекта, а для исследования имеет важное значение объемный или весовой параметр изучаемого явления, то для характеристики выборки целесообразно воспользоваться средней кубической величиной. Определяющее свойство для нее звучит следующим образом: при замене всех значений средним кубическим сумма кубов наблюдений не изменится. Средняя кубическая является степенной средней третьего порядка. Если в (19) подставить значение $m = 3$, то мы получим равенство, выражающее определяющее свойство этой средней:

$$\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} \bar{x}_3^3.$$

Для того чтобы получить формулу для вычисления средней кубической, подставим значение $m = 3$ в выражение (20):

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{n} x_i^3}{n}} = \left(\frac{\sum_{i=1}^{n} x_i^3}{n} \right)^{1/3}. \qquad (29)$$

Формула (29) применяется при вычислении средней кубической для несгруппированного набора данных. В том случае, если выполнялась группировка первичных данных, следует воспользоваться формулой

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{k} f_i \cdot x_i^3}{n}}. \qquad (30)$$

Рассмотрим пример, поясняющий использование средней кубической величины. В табл. 18 приводятся некоторые параметры, характеризующие модельные деревья сосны, и вспомогательные вычисления для рассматриваемого примера.

Таблица 18. **Характеристика модельных деревьев сосны**

Номер модельного дерева	Диаметр на высоте груди (d_m), м^3	Вес стержневого корня дерева V_r, кг	d_m^3
1	2	3	4
1	8,7	4,97	658,503
2	9,6	4,72	884,736
3	18,9	34,12	6751,269
4	15,8	14,99	3944,312
5	15,0	12,79	3341,362
6	19,8	40,3	7821,346
Итого	87,8	111,89	23 401,52

У нас имеются данные, характеризующие линейный размер изучаемого объекта (диаметр на высоте груди). Попробуем оценить среднее значение этого показателя так, чтобы оно адекватно характеризовало объемный (в приводимом примере весовой) признак изучаемого объекта. Для этого наиболее целесообразно воспользоваться средней кубической. Подставляя значение суммы кубов диаметров на высоте груди из 4-й колонки табл. 18 в формулу (29), получим

$$\bar{x}_3 = \sqrt[3]{\frac{23401{,}52}{6}} = 15{,}741\,\text{см}.$$

Как правило, объемные (весовые) показатели пропорциональны третьей степени линейных показателей при условии подобия формы объектов. С некоторой степенью приближения мы можем считать, что форма деревьев подобна у различных сосен. Тогда можно записать соотношение, связывающее вес стержневого корня и диаметр на высоте груди у деревьев сосны:

$$V_r = K \cdot d_m^3, \tag{31}$$

или для совокупности деревьев:

$$\sum V_r = \sum K \cdot d_m^3 = K \cdot \sum d_m^3, \tag{32}$$

где V_r – вес стержневого корня, кг; K – коэффициент пропорциональности, кг/см^3; d_m – диаметр ствола на высоте груди, см. Найдем значение коэффициента K для анализируемой выборки. Подставляя общий вес всех стержневых корней модельных деревьев и сумму кубов диаметров на высоте груди из 2-й и 3-й колонок таблицы 18 в формулу (32) и выразив K, получим

$$K = \frac{\sum V_r}{\sum d_m^3} = \frac{111{,}89}{23\,401{,}52} = 0{,}004\,781\,\text{кг}/\text{см}^3.$$

Пользуясь полученным значением коэффициента $K = 0{,}004\,781\,\text{кг}/\text{см}^3$ и средним кубическим диаметром $\bar{x}_3 = 15{,}741\,\text{см}$, с учетом формулы (31), вычислим, какой вес имеют стержневые корни всех модельных деревьев анализируемой выборки:

$$M_r = n \cdot \bar{V}_r = n \cdot K \cdot \bar{x}_3^3 = 6 \cdot 0{,}004\,781 \cdot 15{,}741^3 = 111{,}88\,\text{кг}.$$

Пользуясь средней кубической, в данном примере нам удалось оценить вес стержневых корней всех модельных деревьев анализируемой выборки с ошибкой 0,01 кг.

Посмотрим, какой точности удалось бы достигнуть, если бы мы использовали для оценки среднего веса стержневого корня среднюю арифметическую величину. Найдем средний арифметический диаметр на высоте груди для анализируемых модельных деревьев, разделив сумму из 2-й колонки табл. 18 на число наблюдений:

$$\bar{x} = \frac{87{,}8}{6} = 14{,}63\,\text{см}.$$

С учетом формулы (31) на основе среднего арифметического значения диаметра на высоте груди и среднего значения коэффициента K получим

$$M_r = n \cdot \overline{V}_r = n \cdot K \cdot \overline{x}^3 = 6 \cdot 0{,}004\,781 \cdot 14{,}633^3 = 89{,}88 \text{ кг}.$$

Полученное значение на 22,01 кг меньше, чем действительное значение общего веса стержневых корней деревьев сосны для анализируемой выборки. Как показывает рассмотренный пример, средняя кубическая оказывается более предпочтительной в том случае, если для целей исследования имеет важное значение объемная или весовая характеристика изучаемого объекта, а учитываемый признак характеризует линейный размер объекта.

Средняя геометрическая. В рассмотренных выше случаях, подставляя в формулу (20) различные значения порядка степенных средних m, мы получали формулы для их вычисления. Такой подход не годится для случая, когда $m = 0$. Попробуем найти предел, к которому стремится степенная средняя, если ее порядок m стремится к нулю:

$$\lim_{m \to 0} \overline{x}_m = \lim_{m \to 0} \left(\frac{\sum\limits_{i=1}^{n} x_i^m}{n} \right)^{1/m}. \tag{33}$$

При $m \to 0$ выражение $\dfrac{\sum\limits_{i=1}^{n} x_i^m}{n} \to 1$, а $1/m \to \infty$. Получаем неопределенность вида 1^∞. Для того чтобы ее раскрыть, введем новую переменную

$$\alpha(m) = \frac{\sum\limits_{i=1}^{n} x_i^m}{n} - 1, \tag{34}$$

которая стремится к нулю при $m \to 0$. Используя (34), выражение (33) можно переписать следующим образом:

$$\lim_{m \to 0} \left(\frac{\sum\limits_{i=1}^{n} x_i^m}{n} \right)^{1/m} = \lim_{m \to 0} \left((1 + \alpha(m))^{1/\alpha(m)} \right)^{\alpha(m)/m}. \tag{35}$$

Так как

$$\lim_{\beta\to 0}(1+\beta)^{1/\beta} = e,$$

то

$$\lim_{m\to 0}(1+\alpha(m))^{1/\alpha(m)} = e. \tag{36}$$

Кроме того,

$$\lim_{m\to 0}\frac{\alpha(m)}{m} = \lim_{m\to 0}\frac{1}{m}\cdot\left(\frac{\sum\limits_{i=1}^{n} x_i^m}{n} - 1\right) = \lim_{m\to 0}\left(\frac{\sum\limits_{i=1}^{n} x_i^m}{n\cdot m} - \frac{1}{m}\right) =$$

$$= \lim_{m\to 0}\left(\frac{\sum\limits_{i=1}^{n} x_i^m - n}{n\cdot m}\right) = \lim_{m\to 0}\left(\frac{\sum\limits_{i=1}^{n}\left(x_i^m - 1\right)}{n\cdot m}\right) = \frac{1}{n}\cdot\lim_{m\to 0}\left(\frac{\sum\limits_{i=1}^{n}\left(x_i^m - 1\right)}{m}\right) =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\lim_{m\to 0}\frac{x_i^m - 1}{m} = \frac{1}{n}\cdot\sum_{i=1}^{n}\ln x_i = \frac{1}{n}\cdot\ln\prod_{i=1}^{n} x_i = \ln\left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Используя (35) и (36), выражение (34) можно переписать в следующем виде:

$$\lim_{m\to 0}\left(\frac{\sum\limits_{i=1}^{n} x_i^m}{n}\right)^{1/m} = \lim_{m\to 0}\left((1+\alpha(m))^{1/\alpha(m)}\right)^{\alpha(m)/m} =$$

$$= e^{\ln\left(\prod\limits_{i=1}^{n} x_i\right)^{1/n}} = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = \sqrt[n]{\prod_{i=1}^{n} x_i}.$$

Такой предельный случай степенной средней m-ного порядка при $m \to 0$ и является средней геометрической величиной. Вычислить эту среднюю в том случае, если исходные данные не подвергались первичной группировке, можно по формуле

$$\bar{x}_g = \sqrt[n]{x_1\cdot x_2\cdot\ldots\cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}. \tag{37}$$

Если экспериментальные данные были предварительно сгруппированы, то среднюю геометрическую можно вычислить, пользуясь формулой

$$\bar{x}_g = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_k^{f_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}}.$$

В данном случае мы не можем сформулировать определяющее свойство по аналогии с рассмотренными выше степенными средними, так как сумма нулевых степеней наблюдений, как, впрочем, и любых других чисел, которыми они могли бы быть заменены, всегда равна числу наблюдений *n*. Однако анализ формулы (37) дает нам основание заключить, что если заменить все значения в выборке средним геометрическим, то произведение наблюдений не изменится, т. е. справедливо равенство:

$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} \bar{x}_g. \tag{38}$$

Действительно, используя (37), правую часть равенства (38) можно преобразовать следующим образом:

$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} \bar{x}_g = \bar{x}_g^n = \left(\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}\right)^n = \left(\sqrt[n]{\prod_{i=1}^{n} x_i}\right)^n = \prod_{i=1}^{n} x_i.$$

Как правило, средняя геометрическая используется в тех случаях, когда анализируются темпы роста признака. Под темпом роста в данном случае понимается отношение величины признака в какой-то момент времени к величине этого же признака в предыдущий момент учета:

$$x_i = \frac{z}{z_{i-1}}, \tag{39}$$

где x_i – темп роста в *i*-тый период времени; z_i – величина признака в *i*-тый момент учета (конец *i*-того периода времени); z_{i-1} – величина признака в *i*-тый момент учета (начало *i*-того периода времени).

Рассмотрим использование средней геометрической на примере анализа темпов роста по высоте древесного ствола. В табл. 16 приведены значения высоты дерева в различных возрастах. На основании этих данных вычислим темпы роста как отношение высоты дерева теперь к высоте дерева 10 лет назад (табл. 19). Средняя геометрическая темпов роста согласно (37) равна

$$\overline{x}_{g} = \sqrt[9]{2{,}889 \cdot 2{,}538 \cdot 1{,}697 \cdot 1{,}214 \cdot 1{,}250 \cdot 1{,}218 \cdot 1{,}150 \cdot 1{,}134 \cdot 1{,}074} =$$

$$= \sqrt[9]{32{,}2116} = 1{,}4708.$$

Таблица 19. **Темпы роста древесного ствола по высоте**

Возраст дерева, лет	Высота дерева, м	Темпы роста дерева
10	0,9	–
20	2,6	2,889
30	6,6	2,538
40	11,2	1,697
50	13,6	1,214
60	17,0	1,250
70	20,7	1,218
80	23,8	1,150
90	27,0	1,134
100	29,0	1,074

Пользуясь средним геометрическим значением темпов роста дерева в высоту, мы можем на основании высоты дерева в начальный момент времени 10 лет (x_{10}) определить высоту дерева в конечный момент времени 100 лет (x_{100}):

$$x_{100} = x_{10} \cdot \overline{x}_g^9 = 0{,}9 \cdot 1{,}4708^9 = 29{,}0 \text{ м.}$$

Как видим, использование средней геометрической величины позволяет нам правильно определить значение признака в конечный момент времени. Для сравнения попробуем решить эту же задачу, пользуясь средней арифметической величиной, которая, согласно (18) для рассматриваемого примера равна

$$\overline{x} = \frac{2{,}889 + 2{,}538 + 1{,}697 + 1{,}214 + 1{,}250 + 1{,}218 + 1{,}150 + 1{,}134 + 1{,}074}{9} =$$

$$= \frac{14{,}164}{9} = 1{,}5738.$$

Высота дерева в 100 лет при использовании средней арифметической величины будет равна

$$x_{100} = x_{10} \cdot \overline{x}^9 = 0{,}9 \cdot 1{,}5738^9 = 0{,}9 \cdot 41{,}5334 = 37{,}4,$$

что значительно превышает истинное значение.

При вычислении средней геометрической по формуле (37) приходится находить произведение довольно большого числа сомножителей и вычислять корень такой же большой степени, что может привести к значительным трудностям. В связи с этим зачастую формулу (37) преобразуют путем логарифмирования левой и правой ее частей и вычисляют логарифм средней геометрической, пользуясь преобразованным вариантом формулы:

$$\log_a \bar{x}_g = \log_a \sqrt[n]{\prod_{i=1}^{n} x_i} = \frac{1}{n}\log_a\left(\prod_{i=1}^{n} x_i\right) = \frac{\sum\limits_{i=1}^{n}\log_a x_i}{n}. \tag{40}$$

Далее путем потенцирования вычисленного логарифма находят среднюю геометрическую величину. Основание логарифма a может выбираться любым, но, как правило, используются десятичные или натуральные логарифмы.

Формулу (40) можно использовать в том случае, если вычисления выполняются на основании несгруппированного набора данных. Если же среднюю геометрическую надо вычислить, пользуясь сгруппированными в статистический ряд данными, то следует воспользоваться формулой

$$\log_a \bar{x}_g = \log_a \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}} = \frac{1}{n}\log_a\left(\prod_{i=1}^{k} x_i^{f_i}\right) = \frac{\sum\limits_{i=1}^{k}\log_a x_i^{f_i}}{n} = \\ = \frac{\sum\limits_{i=1}^{k} f_i \cdot \log_a x_i}{n}. \tag{41}$$

В том случае, если средняя геометрическая величина определяется для темпов роста какого-либо признака, вычисленных для последовательных равных периодов времени, можно воспользоваться более простой формулой:

$$\bar{x}_g = \sqrt[n]{\frac{z_n}{z_0}} \quad \text{или} \quad \log_a \bar{x}_g = \frac{\log_a z_n - \log_a z_0}{n},$$

где z_0 – значение признака в начальный момент учета; z_n – значение признака в конечный момент учета. Действительно, подставляя (39) в (37), получаем

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\frac{z_1}{z_0} \cdot \frac{z_2}{z_1} \cdot \ldots \cdot \frac{z_n}{z_{n-1}}} = \sqrt[n]{\frac{z_n}{z_0}}.$$

Рассмотрим порядок вычисления средних величин на примере вариационных рядов по диаметрам и высотам (табл. 12 и 13). Для выполнения вычислений составим вспомогательную табл. 20.

Теперь, подставив в формулы (21) и (28) соответствующие суммы из 2-й, 3-й и 4-й колонок табл. 20, вычислим среднее арифметическое значение:

$$\bar{x} = \frac{\sum_{i=1}^{k}(x_i \cdot f_i)}{n} = \frac{6320{,}40}{200} = 31{,}60 \text{ см} \qquad (42)$$

и среднее квадратическое:

$$\bar{x}_2 = \sqrt[2]{\frac{\sum_{i=1}^{n}\left(x_i^2 \cdot f_i\right)}{n}} = \sqrt[2]{\frac{210\,797{,}2}{200}} = 32{,}47 \text{ см.}$$

Таблица 20. **Вычисление средних значений (диаметры)**

x_i	f_i	$f_i x_i$	$f_i x_i^2$	$f_i x_i^3$	f_i/x_i	$ln(x_i)$	$f_i ln(x_i)$
1	2	3	4	5	6	7	8
17,65	3	52,95	934,6	16 495,7	0,1700	2,871	8,613
20,85	11	229,35	4781,9	99 702,6	0,5276	3,037	33,407
24,05	29	697,45	16 773,7	403 407,5	1,2058	3,180	92,220
27,25	39	1062,75	28 959,9	789 157,3	1,4312	3,305	128,895
30,45	32	974,40	29 670,5	903 466,7	1,0509	3,416	109,312
33,65	33	1110,45	37 366,6	1 257 386,1	0,9807	3,516	116,028
36,85	23	847,55	31 232,2	1 150 906,6	0,6242	3,607	82,961
40,05	10	400,50	16 040,0	642 402,0	0,2497	3,690	36,900
43,25	9	389,25	16 835,1	728 118,1	0,2081	3,767	33,903
46,45	3	139,35	6472,8	300 661,6	0,0646	3,838	11,514
49,65	4	198,60	9860,5	489 573,8	0,0806	3,905	15,620
52,85	2	105,70	5586,2	295 230,7	0,0378	3,967	7,934
56,05	2	112,10	6283,2	352 173,4	0,0357	4,026	8,052
Сумма	200	6320,40	210 797,2	7 428 682,1	6,6669		685,359

Для вычисления среднего кубического в соответствии с формулой (30) следует сумму из 5-й колонки табл. 20 разделить на объем выборки:

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{n}\left(x_i^3 \cdot f_i\right)}{n}} = \sqrt[3]{\frac{7\,428\,682{,}1}{200}} = 33{,}37 \text{ см.}$$

Воспользовавшись формулой (27) и данными из табл. 20, вычислим среднюю гармоническую величину:

$$\bar{x}_{-1} = \frac{n}{\sum_{i=1}^{k}(f_i/x_i)} = \frac{200}{6{,}6669} = 30{,}00 \text{ см.}$$

Для вычисления средней геометрической сначала вычислим ее логарифм по формуле (41):

$$\ln \bar{x}_g = \frac{\sum_{i=1}^{k} f_i \cdot \ln x_i}{n} = \frac{685{,}359}{200} = 3{,}4268,$$

а затем потенцированием найдем среднюю геометрическую:

$$\bar{x}_g = e^{\ln \bar{x}_g} = e^{3{,}4268} = 30{,}78 \text{ м}^2.$$

Аналогичным образом определим средние значения для ряда распределения высот деревьев в древостое. Сначала составим вспомогательную табл. 21.

Таблица 21. **Вычисление средних значений (высоты)**

x_i	f_i	$f_i x_i$	$f_i x_i^2$	$f_i x_i^3$	f_i/x_i	$ln(x_i)$	$f_i ln(x_i)$
17,05	2	34,10	581,41	9913,04	0,1173	2,836	5,672
18,05	1	18,05	325,80	5880,69	0,0554	2,893	2,893
19,05	4	76,20	1451,61	27 653,17	0,2100	2,947	11,788
20,05	5	100,25	2010,01	40 300,70	0,2494	2,998	14,990
21,05	2	42,10	886,21	18 654,72	0,0950	3,047	6,094
22,05	13	286,65	6320,63	139 369,90	0,5896	3,093	40,209
23,05	25	576,25	13 282,56	306 163,00	1,0846	3,138	78,450
24,05	31	745,55	17 930,48	431 228,00	1,2890	3,180	98,580
25,05	32	801,60	20 080,08	503 006,00	1,2774	3,221	103,072
26,05	43	1120,15	29 179,91	760 136,70	1,6507	3,260	140,180
27,05	24	649,20	17 560,86	475 021,30	0,8872	3,298	79,152
28,05	14	392,70	11 015,24	308 977,50	0,4991	3,334	46,676
29,05	3	87,15	2531,71	73 546,18	0,1033	3,369	10,107
30,05	1	30,05	903,00	27 135,15	0,0333	3,403	3,403
Сумма	200	4960,00	124 059,51	3 126 986,05	8,1413		641,266

Затем, подставляя суммы из табл. 21 в формулы (21), (27), (28), (30) и (41), вычислим степенные средние:

средняя арифметическая:

$$\bar{x} = \frac{\sum_{i=1}^{k}(x_i \cdot f_i)}{n} = \frac{4960,00}{200} = 24,80 \text{ м}; \quad (43)$$

средняя квадратическая:

$$\bar{x}_2 = \sqrt[2]{\frac{\sum_{i=1}^{n}\left(x_i^2 \cdot f_i\right)}{n}} = \sqrt[2]{\frac{124\,059,51}{200}} = 24,91 \text{ м};$$

средняя кубическая:

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{n}\left(x_i^3 \cdot f_i\right)}{n}} = \sqrt[3]{\frac{3\,126\,986,05}{200}} = 25,01 \text{ м};$$

средняя гармоническая:

$$\bar{x}_{-1} = \frac{n}{\sum_{i=1}^{k}(f_i / x_i)} = \frac{200}{8,1413} = 24,57 \text{ м};$$

средняя геометрическая:

$$\ln \bar{x}_g = \frac{\sum_{i=1}^{k} f_i \cdot \ln x_i}{n} = \frac{641,266}{200} = 3,2063;$$

$$\bar{x}_g = e^{\ln \bar{x}_g} = e^{3,2063} = 24,69 \text{ м}^2.$$

2.2. Показатели вариации

Средние величины указывают на то значение признака, вокруг которого группируются анализируемые наблюдения. Однако вокруг одного и того же значения признака наблюдения могут располагаться совершенно по-разному. К примеру, они все могут очень незначительно отличаться от среднего значения, располагаясь плотной группой вокруг него, или, напротив, иметь сильный разброс. Для того чтобы отразить характер расположения наблюдений вокруг среднего, и служат показатели вариации. Рассмотрим некоторые из них.

Размах вариации. Наиболее простым показателем, характеризующим распределение вариант вокруг среднего, является размах вариации, который вычисляется как разность между максимальным и минимальным значениями признака, называемыми в биометрии также *лимитами* (от латинского слова limes - предел) и обозначаемыми символом lim:

$$R = x_{\max} - x_{\min}. \tag{44}$$

Если наблюдения плотно группируются вокруг среднего, то лимиты располагаются близко друг к другу, и размах вариации оказывается небольшим. Если же разброс данных сильный, то, как правило, минимальная и максимальная варианты располагаются далеко друг от друга, и размах вариации получается большим.

Размах вариации является насколько простым показателем, настолько же и ненадежным. Дело в том, что он вычисляется на основании значений лимитов, а последние, в свою очередь, являются очень неустойчивыми статистиками и могут значительно варьировать от выборки к выборке. Кроме того, так как при вычислении размаха вариации используются только две крайние варианты, он не дает нам никакой информации о характере распределения всех остальных вариант, располагающихся ближе к среднему.

Среднее линейное отклонение. Этих недостатков лишено среднее линейное отклонение, которое вычисляется как средняя арифметическая величина из абсолютных значений отклонений всех наблюдений от их средней арифметической:

$$\bar{d} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}.$$

В том случае, если среднее линейное отклонение вычисляется на основе сгруппированных данных, следует воспользоваться формулой

$$\bar{d} = \frac{\sum_{i=1}^{k} f_i \cdot |x_i - \bar{x}|}{n}. \tag{45}$$

В данной статистике применяются именно абсолютные величины отклонений, так как среднее значение самих отклонений не может служить мерой вариации признака. Дело в том, что согласно свойству 1 средней арифметической величины, сумма отклонений вариант от средней арифметической равна нулю.

Среднее линейное отклонение характеризует изменчивость показателя гораздо лучше, чем размах вариации. Тем не менее, наиболее широкое применение получил другой показатель, основанный на квадратах отклонений.

Эмпирическая дисперсия - это средний квадрат отклонений вариант от среднего арифметического. Данный показатель получил свое название от латинского слова dispersio - рассеяние. Вычислить эту статистику можно по формуле

$$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \tag{46}$$

или, если речь идет о сгруппированном наборе данных, по формуле

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n}. \tag{47}$$

Выборочная дисперсия, рассчитанная по формуле (46), дает смещенную оценку генеральной дисперсии. Для того чтобы получить несмещенную оценку, в формулу необходимо добавить сомножитель $\frac{n}{n-1}$, называемый *поправкой Бесселя*:

$$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{48}$$

или

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n-1} \tag{49}$$

для сгруппированного набора данных. Величина $n-1$ из формул (48) и (49) называется *числом степеней свободы*. Она показывает, сколько в данном случае имеется независимых наблюдений. Дело в том, что в формуле (46) используется средняя арифметическая величина, вычисленная по данным той же самой выборки по формуле (18). В связи с этим независимыми наблюдениями в данной выборке можно считать только $n-1$ элементов, так как последний n-ный элемент полностью определяется остальными $n-1$ элементами и средней арифметической.

Эмпирическая дисперсия обладает рядом свойств.

1. Свойство *минимальности*, которое заключается в том, что эмпирическая дисперсия меньше среднего квадратов отклонений

наблюдений от любой точки, не равной средней арифметической, т. е.

$$S_x^2 < \frac{\sum_{i=1}^{n}(x_i - a)^2}{n}, \text{ если } a \neq \overline{x}.$$

Действительно, согласно свойству 4 средней арифметической, сумма квадратов отклонений вариант от их средней арифметической меньше, чем сумма квадратов отклонений от любой другой величины, не равной средней. Если обе части неравенства (23), выражающего это свойство, разделить на объем выборки n, получим

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} < \frac{\sum_{i=1}^{n}(x_i - a)^2}{n}, \text{ для любого } a \neq \overline{x}. \tag{50}$$

Учитывая (46), выражение (50) можно преобразовать следующим образом:

$$S_x^2 < \frac{\sum_{i=1}^{n}(x_i - a)^2}{n}, \text{ если } a \neq \overline{x}.$$

2. Дисперсия постоянной величины равна нулю. В самом деле, если все элементы выборки объема n равны между собой и равны величине c, то эмпирическая дисперсия будет равна

$$S_c^2 = \frac{\sum_{i=1}^{n}(c - \overline{c})^2}{n}. \tag{51}$$

Так как

$$\overline{c} = \frac{\sum_{i=1}^{n} c}{n} = \frac{n \cdot c}{n} = c,$$

то (51) можно преобразовать следующим образом:

$$S_c^2 = \frac{\sum_{i=1}^{n}(c - \overline{c})^2}{n} = \frac{\sum_{i=1}^{n}(c - c)^2}{n} = \frac{\sum_{i=1}^{n} 0}{n} = 0.$$

3. Если к каждому наблюдению в выборке прибавить константу c ($c \in [-\infty, +\infty]$), то выборочная дисперсия не изменится. Действительно, учитывая свойство 2 среднего арифметического (22) и (46), можем записать

$$S_{x+c}^2 = \frac{\sum_{i=1}^{n}\left((x_i + c) - \left(\overline{x+c}\right)\right)^2}{n} = \frac{\sum_{i=1}^{n}\left((x_i + c) - \left(\bar{x} + c\right)\right)^2}{n} =$$

$$= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = S_x^2.$$

4. Если каждое наблюдение в выборке умножить на константу c ($c \in [-\infty, +\infty]$), то эмпирическая дисперсия увеличится в c^2 раз, т. е.

$$S_{c \cdot x}^2 = c^2 \cdot S_x^2.$$

Покажем это. Учитывая свойство 3 среднего арифметического и выражение (46), можно записать

$$S_{x \cdot c}^2 = \frac{\sum_{i=1}^{n}\left(c \cdot x_i - \overline{cx}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(c \cdot x_i - c \cdot \bar{x}\right)^2}{n} =$$

$$= \frac{\sum_{i=1}^{n} c^2 \cdot \left(x_i - \bar{x}\right)^2}{n} = c^2 \cdot \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = c^2 \cdot S_x^2.$$

5. Если ряд наблюдений состоит из двух групп, то выборочная дисперсия такого ряда равна сумме средневзвешенной из дисперсий групп и средневзвешенной из квадратов отклонений групповых средних от общей средней. При этом в качестве весов служат объемы выборок. Это свойство можно записать в виде формулы

$$S^2 = \frac{S_1^2 \cdot n_1 + S_2^2 \cdot n_2}{n_1 + n_2} + \frac{\left(\bar{x}_1 - \bar{x}\right)^2 \cdot n_1 + \left(\bar{x}_2 - \bar{x}\right)^2 \cdot n_2}{n_1 + n_2}, \qquad (52)$$

где S^2 – дисперсия для всего ряда; $\bar{x}$ – средняя арифметическая для всего ряда; S_1^2 – дисперсия для 1-й группы наблюдений; $\bar{x}_1$ – средняя арифметическая для 1-й группы наблюдений; n_1 – число наблюдений в 1-й группе наблюдений; S_2^2 – дисперсия для 2-й группы наблюдений; $\bar{x}_2$ – средняя арифметическая для 2-й группы наблюдений; n_2 – число наблюдений во 2-й группе наблюдений. Докажем это. Для начала покажем, что

$$S^2 = \overline{x^2} - \left(\bar{x}\right)^2. \qquad (53)$$

Учитывая (18) и (46), выражение (53) можно преобразовать следующим образом:

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2 - 2\cdot\sum_{i=1}^{n} x_i \cdot \bar{x} + \sum_{i=1}^{n}\left(\bar{x}\right)^2}{n} =$$

$$\frac{\sum_{i=1}^{n} x_i^2}{n} - 2\cdot\bar{x}\cdot\frac{\sum_{i=1}^{n} x_i}{n} + \frac{n\cdot\left(\bar{x}\right)^2}{n} = \overline{x^2} - 2\cdot\left(\bar{x}\right)^2 + \left(\bar{x}\right)^2 = \overline{x^2} - \left(\bar{x}\right)^2,$$

что и требовалось доказать.

Используя (53) и (18), первое слагаемое в выражении (52) можно преобразовать следующим образом:

$$\frac{S_1^2\cdot n_1 + S_2^2\cdot n_2}{n_1+n_2} = \frac{\overline{x_1^2}\cdot n_1 + \overline{x_2^2}\cdot n_2 - \left(\overline{x_1}\right)^2\cdot n_1 - \left(\overline{x_2}\right)^2\cdot n_2}{n_1+n_2} =$$

$$= \frac{\sum_{i=1}^{n_1} x_i^2 + \sum_{i=n_1+1}^{n_1+n_2} x_i^2 - \left(\overline{x_1}\right)^2\cdot n_1 - \left(\overline{x_2}\right)^2\cdot n_2}{n_1+n_2} =$$

$$= \frac{\sum_{i=1}^{n_1+n_2} x_i^2 - \left(\overline{x_1}\right)^2\cdot n_1 - \left(\overline{x_2}\right)^2\cdot n_2}{n_1+n_2}.$$

Выражение (25) (5 свойство среднего арифметического) можно преобразовать к виду

$$\bar{x}\cdot(n_1+n_2) = \overline{x_1}\cdot n_1 + \overline{x_2}\cdot n_2. \qquad (54)$$

Далее, используя (54), преобразуем второе слагаемое из выражения (52) следующим образом:

$$\frac{\left(\overline{x_1}-\bar{x}\right)^2\cdot n_1 + \left(\overline{x_2}-\bar{x}\right)^2\cdot n_2}{n_1+n_2} =$$

$$= \frac{\left(\overline{x_1}\right)^2\cdot n_1 - 2\cdot\overline{x_1}\cdot\bar{x}\cdot n_1 + \left(\bar{x}\right)^2\cdot n_1 + \left(\overline{x_2}\right)^2\cdot n_2 - 2\cdot\overline{x_2}\cdot\bar{x}\cdot n_2 + \left(\bar{x}\right)^2\cdot n_2}{n_1+n_2} =$$

$$= \frac{\left(\overline{x_1}\right)^2\cdot n_1 + \left(\overline{x_2}\right)^2\cdot n_2 + \left(\bar{x}\right)^2\cdot(n_1+n_2) - 2\cdot\bar{x}\cdot(\overline{x_1}\cdot n_1 + \overline{x_2}\cdot n_2)}{n_1+n_2} =$$

$$= \frac{\left(\overline{x_1}\right)^2\cdot n_1 + \left(\overline{x_2}\right)^2\cdot n_2 + \left(\bar{x}\right)^2\cdot(n_1+n_2) - 2\cdot\left(\bar{x}\right)^2\cdot(n_1+n_2)}{n_1+n_2} =$$

$$= \frac{\left(\overline{x_1}\right)^2 \cdot n_1 + \left(\overline{x_2}\right)^2 \cdot n_2 - \left(\overline{x}\right)^2 \cdot (n_1 + n_2)}{n_1 + n_2}.$$

Подставляя найденные выражения для слагаемых в выражение (52) и учитывая (53), получим

$$S^2 = \frac{S_1^2 \cdot n_1 + S_2^2 \cdot n_2}{n_1 + n_2} + \frac{\left(\overline{x_1} - \overline{x}\right)^2 \cdot n_1 + \left(\overline{x_2} - \overline{x}\right)^2 \cdot n_2}{n_1 + n_2} =$$

$$= \frac{\sum_{i=1}^{n_1+n_2} x_i^2 - \left(\overline{x_1}\right)^2 \cdot n_1 - \left(\overline{x_2}\right)^2 \cdot n_2 + \left(\overline{x_1}\right)^2 \cdot n_1 + \left(\overline{x_2}\right)^2 \cdot n_2 - \left(\overline{x}\right)^2 \cdot (n_1 + n_2)}{n_1 + n_2} =$$

$$= \frac{\sum_{i=1}^{n_1+n_2} x_1^2}{n_1 + n_2} - \left(\overline{x}\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2 = S^2,$$

что и требовалось доказать.

Данное свойство можно обобщить на случай, когда выборка состоит из k групп наблюдений. Чтобы сделать такое обобщение, целесообразно ввести понятия межгрупповой и внутригрупповой дисперсии.

Межгрупповой дисперсией мы будем называть среднюю арифметическую квадратов отклонений групповых средних $\overline{x}_i$ от средней для всей выборки $\overline{x}$ (причем в качестве весов при вычислении этой дисперсии используются объемы групп n_i). Эту дисперсию можно вычислить по формуле

$$\delta^2 = \frac{\sum_{i=1}^{k} n_i \cdot \left(\overline{x_i} - \overline{x}\right)^2}{\sum_{i=1}^{k} n_i}. \qquad (55)$$

Внутригрупповой дисперсией, или *средней групповых дисперсий*, $\overline{S_i^2}$ мы будем называть среднюю арифметическую групповых дисперсий S_i^2. В качестве весов в данном случае используются объемы групп n_i:

$$\overline{S_i^2} = \frac{\sum_{i=1}^{k} n_i \cdot S_i^2}{\sum_{i=1}^{k} n_i}. \tag{56}$$

С использованием понятий межгрупповой и внутригрупповой дисперсий данное свойство в случае, когда выборка состоит из K групп наблюдений, будет звучать следующим образом: дисперсия выборки, состоящей из K групп наблюдений, равна сумме внутригрупповой и межгрупповой дисперсий, т. е.

$$S^2 = \overline{S_i^2} + \delta^2. \tag{57}$$

Среднеквадратическое отклонение. Дисперсия является очень важной характеристикой выборки, однако в некоторых случаях для характеристики изменчивости анализируемых данных удобнее использовать среднеквадратическое отклонение, которое является квадратным корнем из дисперсии:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \text{ – смещенная оценка;}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \text{ – несмещенная оценка.} \tag{58}$$

Или для сгруппированного набора данных:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n}} \text{ – смещенная оценка;} \tag{59}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n-1}} \text{ – несмещенная оценка.} \tag{60}$$

В отличие от дисперсии, среднеквадратическое отклонение выражается в тех же единицах измерения, что и анализируемый признак. В связи с этим данный показатель является более естественным и легче поддается анализу.

Коэффициент вариации. Дисперсия и среднеквадратическое отклонение довольно полно характеризуют вариацию признака, однако в некоторых случаях удобнее иметь показатель, который оценивает разброс данных не в абсолютных величинах, а в относительных. Таким показателем является коэффициент вариации, который показывает, сколько процентов составляет среднеквадратическое отклонение от среднего арифметического:

$$V = \frac{S_x}{\bar{x}} \cdot 100\%. \tag{61}$$

В биометрии этот показатель часто оказывается весьма полезным. Дело в том, что анализу подвергаются, как правило, объекты живой природы, а они с течением времени изменяют свои размеры: растут. В связи с этим часто необходимо анализировать выборки, сделанные для объектов с разным средним возрастом, а следовательно, и разными средними размерами. Если в таких случаях необходимо сравнить степень изменчивости признака в разных выборках, то удобнее оперировать коэффициентом вариации, так как он дает нам величину вариации по отношению к среднему значению.

Коэффициент асимметрии. Рассмотренные выше показатели довольно полно характеризуют анализируемые признаки, однако ни один из них не отражает степень симметричности распределения наблюдений относительно среднего значения. Ведь на практике довольно часто отклонения признака от среднего арифметического в меньшую и большую стороны носят неодинаковый характер.

Для того чтобы оценить степень такой неравномерности распределения наблюдений относительно среднего арифметического, используют коэффициент асимметрии, который можно вычислить по формуле

$$A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot S_x^3},$$

или для сгруппированного набора данных:

$$A = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^3}{n \cdot S_x^3}. \tag{62}$$

Коэффициент асимметрии может принимать как положительные, так и отрицательные значения. В том случае, если левая ветвь распределения более пологая и длинная, а вершина кривой смещена вправо относительно среднего арифметического, то коэффициент асимметрии для такого распределения имеет отрицательное значение. Такая асимметрия называется левосторонней, или отрицательной (рис. 8).

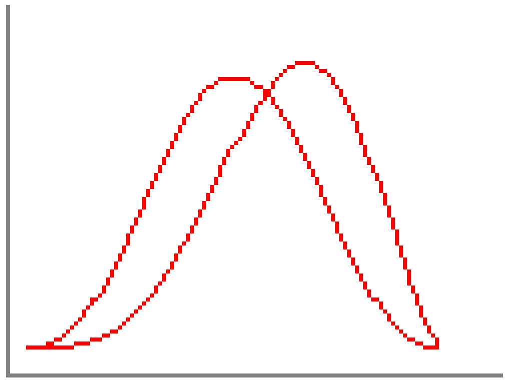


Рис. 8. Левосторонняя, или отрицательная, асимметрия кривой распределения

Если распределение имеет более длинную и пологую правую ветвь, а его вершина смещена влево относительно среднего арифметического, то в таком случае имеет место правосторонняя, или положительная, асимметрия (рис. 9). Коэффициент асимметрии в таком случае будет положительным.

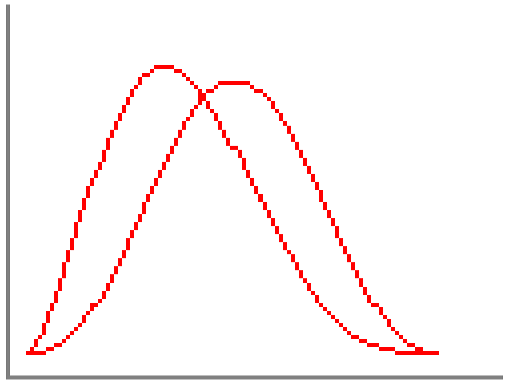


Рис. 9. Правосторонняя, или положительная, асимметрия кривой распределения

Эмпирический коэффициент эксцесса. Кроме того, что распределения наблюдений могут отличаться друг от друга по степени асимметричности, они могут иметь разную крутизну: быть островершинными и плосковершинными. В случае островершинной кривой, когда большое число наблюдений группируется в непосредственной близости от центра распределения, говорят о наличии положительного эксцесса. Кривая распределения имеет отрицательный эксцесс, если она является плосковершинной. Для оценки степени крутизны кривой распределения используется коэффициент эксцесса, который вычисляется по формуле

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \cdot S_x^4} - 3,$$

или для сгруппированного набора данных:

$$E = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4}{n \cdot S_x^4} - 3. \qquad (63)$$

Этот коэффициент построен таким образом, что его значение для нормального распределения[1], как для наиболее изученного и часто используемого, равен нулю. В том случае, если коэффициент эксцесса принимает положительное значение (положительный эксцесс), распределение вариант будет более крутым, чем нормальное распределение (рис. 10).

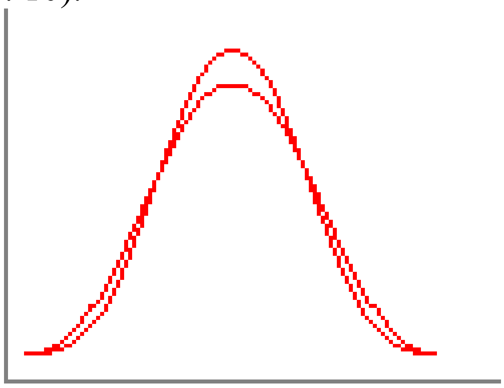


Рис. 10. Положительный эксцесс, островершинная кривая

Когда этот показатель меньше нуля (отрицательный эксцесс), наблюдения будут образовывать более плосковершинную кривую, чем нормальное распределение (рис. 11).

---

[1] Нормальное распределение будет рассматриваться в следующей главе.

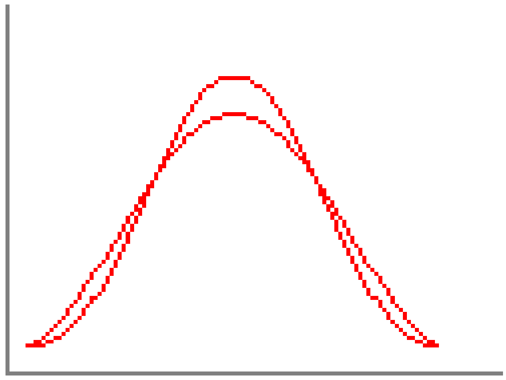


Рис. 11. Отрицательный эксцесс, плосковершинная кривая

Эмпирические моменты. Кроме перечисленных выше показателей, для характеристики эмпирических распределений используется система статистик, называемых моментами. Эмпирический момент порядка q определяется как средняя арифметическая величина из q-тых степеней наблюдаемых значений признака:

$$m_q = \frac{\sum_{i=1}^{n} x_i^q}{n}, \tag{64}$$

или в случае сгруппированного набора данных:

$$m_q = \frac{\sum_{i=1}^{k} f_i \cdot x_i^q}{n}.$$

Такие моменты иногда называют начальными.

Эмпирический начальный момент нулевого порядка всегда равен 1:

$$m_0 = \frac{\sum_{i=1}^{n} x_i^0}{n} = \frac{\sum_{i=1}^{n} 1}{n} = \frac{n}{n} = 1.$$

Эмпирический начальный момент первого порядка – это не что иное, как средняя арифметическая величина. Действительно, если в формулу (64) подставить значение $q = 1$, то мы получим формулу, выражающую среднее арифметическое значение:

$$m_1 = \frac{\sum_{i=1}^{n} x_i^1}{n} = \frac{\sum_{i=1}^{n} x_1}{n} = \bar{x}.$$

Если c – константа, то выражением

$$b_i = \frac{\sum_{i=1}^{n} (x_i - c)^q}{n} \qquad (65)$$

задается момент относительно точки c порядка q.

В случае, когда $c = 0$, получаем начальные моменты. Моменты относительно средней арифметической ($c = \bar{x}$) называются центральными:

$$\mu_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^q}{n},$$

или для сгруппированного набора данных:

$$\mu_k = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^q}{n}. \qquad (66)$$

В силу свойства 1 средней арифметической величины, гласящего, что сумма отклонений вариант от средней арифметической величины равна нулю, эмпирический центральный момент первого порядка также равен нулю:

$$\mu_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n} = 0.$$

Эмпирический центральный момент второго порядка – это не что иное, как выборочная дисперсия:

$$\mu_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = S_x^2.$$

Эмпирические центральные моменты всех порядков обладают двумя свойствами.

1. Если все наблюдения в анализируемой выборке изменить на постоянную величину c $(c \in [-\infty; +\infty])$, то центральный момент q-того порядка не изменится. Действительно, учитывая свойство 2 средней

арифметической величины $(\overline{x+c} = \bar{x} + c)$, имеем:

$$\mu_{q,x-c} = \frac{\sum_{i=1}^{n}((x-c)-\overline{x-c})^q}{n} = \frac{\sum_{i=1}^{n}(x-c-\bar{x}+c)^q}{n} =$$

$$= \frac{\sum_{i=1}^{n}(x-\bar{x})^q}{n} = \mu_{q,x}.$$

2. Если все наблюдения в анализируемой выборке умножить на константу c $(c \in [-\infty;+\infty])$, то центральный момент q-того порядка увеличится в c^q раз. Действительно, учитывая свойство 3 средней арифметической $(\overline{x \cdot c} = \bar{x} \cdot c)$, получим

$$\mu_{q,x\cdot c} = \frac{\sum_{i=1}^{n}((x\cdot c)-\overline{x\cdot c})^q}{n} = \frac{\sum_{i=1}^{n}(x\cdot c-\bar{x}\cdot c)^q}{n} =$$

$$= c^q \cdot \frac{\sum_{i=1}^{n}(x-\bar{x})^q}{n} = c^q \cdot \mu_{q,x}.$$

Если известны моменты q--того порядка и ниже относительно точки a, то на основании этих величин можно определить момент q-того порядка относительно любой другой точки c.

С учетом формулы бинома имеем

$$(x-c)^q = ((x-a)+(a-c))^q = \sum_{j=0}^{q} C_q^j (x-a)^{q-j} \cdot (a-c)^j, \qquad (67)$$

где C_q^j – биномиальный коэффициент:

$$C_q^j = \begin{cases} \dfrac{q!}{j!\cdot(q-j)!}, & 0 \le j \le q, \\ 0, & 0 \le q < j. \end{cases}$$

Используя выражение (67), получим

$$b_q(c) = \frac{\sum_{i=1}^{n}(x-c)^q}{n} = \frac{\sum_{i=1}^{n} C_q^j \cdot (x-a)^{q-j} \cdot (a-c)^j}{n} =$$

$$= \sum_{j=0}^{q} C_q^j \cdot (a-c)^j \cdot \frac{\sum_{i=1}^{n} (x-a)^{q-j}}{n} = \sum_{j=0}^{q} C_q^j \cdot (a-c)^j \cdot b_{q-j}(a). \qquad (68)$$

Для первых четырех моментов выражение (68) выглядит следующим образом:

$$b_1(c) = (a-c) + b_1(a);$$

$$b_2(c) = (a-c)^2 + 2 \cdot (a-c) \cdot b_1(a) + b_2(a);$$

$$b_3(c) = (a-c)^3 + 3 \cdot (a-c)^2 \cdot b_1(a) + 3 \cdot (a-c) \cdot b_2(a) + b_3(a);$$

$$b_4(c) = (a-c)^4 + 4 \cdot (a-c)^3 \cdot b_1(a) + 6 \cdot (a-c)^2 \cdot b_2(a) +$$
$$+ 4 \cdot (a-c) \cdot b_3(a) + b_4(a).$$

В том случае, если величина c равна среднему арифметическому, и учитывая, что

$$a - \bar{x} = \frac{n \cdot a - \sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} a - \sum_{i=1}^{n} x_i}{n} = -\frac{\sum_{i=1}^{n} (x_i - a)}{n} = -b_1(a),$$

получаем

$$\mu_1 = 0;$$

$$\mu_2 = b_2(a) - \left(b_1(a)\right)^2;$$

$$\mu_3 = b_3(a) - 3 \cdot b_1(a) \cdot b_2(a) + 2 \cdot \left(b_1(a)\right)^3;$$

$$\mu_4 = b_4(a) - 4 \cdot b_1(a) \cdot b_3(a) + 6 \cdot \left(b_1(a)\right)^2 \cdot b_2(a) - 3 \cdot \left(b_1(a)\right)^4.$$

Вычислим рассмотренные выше показатели вариации для диаметров и высот 200 деревьев, приведенных в последних четырех столбцах табл. 1 приложения с номерами 17 ,18, 19 и 20.

Проще всего определить размах вариации. Для этого достаточно найти минимальное и максимальное значения и подставить их в формулу (44):

$R = x_{\max} - x_{\min} = 56{,}6 - 17{,}7 = 38{,}9$ – для диаметров;

$R = x_{\max} - x_{\min} = 29{,}6 - 17{,}5 = 12{,}1$ – для высот.

Для того чтобы определить остальные показатели вариации, составим по данным вариационных рядов диаметров и высот вспомогательные табл. 22 и 23.

Таблица 22. **Вычисление показателей вариации (диаметры)**

x_i	f_i	$x_i - \bar{x}$	$f_i \cdot \lvert x_i - \bar{x} \rvert$	$(x_i - \bar{x})^2 \cdot f_i$	$(x_i - \bar{x})^3 \cdot f_i$	$(x_i - \bar{x})^4 \cdot f_i$
17,65	3	–13,95	41,85	583,81	–8144,1	113 610,2
20,85	11	–10,75	118,25	1271,19	–13 665,3	146 902,0
24,05	29	–7,55	218,95	1653,07	–12 480,7	94 229,3
27,25	39	–4,35	169,65	737,98	–3210,2	13 964,4
30,45	32	–1,15	36,80	42,32	–48,7	56,0
33,65	33	2,05	67,65	138,68	284,3	582,8
36,85	23	5,25	120,75	633,94	3328,2	17 473,1
40,05	10	8,45	84,50	714,03	6033,6	50 983,9
43,25	9	11,65	104,85	1221,50	14 230,5	165 785,3
46,45	3	14,85	44,55	661,57	9824,3	145 890,9
49,65	4	18,05	72,20	1303,21	23 522,9	424 588,3
52,85	2	21,25	42,50	903,13	19 191,5	407 819,4
56,05	2	24,45	48,90	1195,61	29 232,7	714 739,5
Сумма	200		1171,40	11 060,04	68 099,0	2 296 625,1

Таблица 23. **Вычисление показателей вариации (высоты)**

x_i	f_i	$x_i - \bar{x}$	$f_i \cdot \lvert x_i - \bar{x} \rvert$	$(x_i - \bar{x})^2 \cdot f_i$	$(x_i - \bar{x})^3 \cdot f_i$	$(x_i - \bar{x})^4 \cdot f_i$
17,05	2	–7,75	15,50	120,13	–931,0	7215,3
18,05	1	–6,75	6,75	45,56	–307,5	2075,6
19,05	4	–5,75	23,00	132,25	–760,4	4372,3
20,05	5	–4,75	23,75	112,81	–535,8	2545,1
21,05	2	–3,75	7,50	28,13	–105,5	395,6
22,05	13	–2,75	35,75	98,31	–270,4	743,6
23,05	25	–1,75	43,75	76,56	–134,0	234,5
24,05	31	–0,75	23,25	17,44	–13,1	9,8
25,05	32	0,25	8,00	2,00	0,5	0,1
26,05	43	1,25	53,75	67,19	84,0	105,0
27,05	24	2,25	54,00	121,50	273,4	615,2
28,05	14	3,25	45,50	147,88	480,6	1562,0
29,05	3	4,25	12,75	54,19	230,3	978,8
30,05	1	5,25	5,25	27,56	144,7	759,7
Сумма	200		358,50	1051,51	–1844,2	21 612,6

Теперь, пользуясь суммами, приведенными в этих таблицах, и формулами (45), (47), (49), (59), (60), (61), (62), (63) и (66), получим оценки остальных показателей вариации для диаметров и высот:

диаметры

$$\overline{d} = \frac{\sum_{i=1}^{k} f_i \cdot |x_i - \overline{x}|}{n} = \frac{1171,40}{200} = 5,857;$$

смещенная оценка дисперсии:

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^2}{n} = \frac{11\,060,04}{200} = 55,30;$$

несмещенная оценка дисперсии:

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^2}{n-1} = \frac{11\,060,04}{199} = 55,58;$$

смещенная оценка среднеквадратического отклонения:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^2}{n}} = \sqrt{\frac{11\,060,04}{200}} = 7,436;$$

несмещенная оценка среднеквадратического отклонения:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{11\,060,04}{199}} = 7,455; \qquad (69)$$

$$V = \frac{S_x}{\overline{x}} \cdot 100\% = \frac{7,455}{31,60} \cdot 100\% = 23,59\%;$$

$$A = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^3}{n \cdot S_x^3} = \frac{68\,099,0}{200 \cdot 7,455^3} = 0,8218; \qquad (70)$$

$$E = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^4}{n \cdot S_x^4} - 3 = \frac{2\,296\,625,1}{200 \cdot 7,455^4} = 3,7177 - 3 = 0,7177; \qquad (71)$$

$$\mu_0 = 1;\ \mu_1 = 0;\ \mu_2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^2}{n} = \frac{11\,059,52}{200} = 55,30;$$

$$\mu_3 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x})^3}{n} = \frac{68\,099,0}{200} = 340,50;$$

$$\mu_4 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4}{n} = \frac{2\,296\,625{,}1}{200} = 11\,483{,}1;$$

высоты

$$\bar{d} = \frac{\sum_{i=1}^{k} f_i \cdot |x_i - \bar{x}|}{n} = \frac{358{,}50}{200} = 1{,}792;$$

смещенная оценка дисперсии:

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n} = \frac{1051{,}51}{200} = 5{,}258;$$

несмещенная оценка дисперсии:

$$S_x^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n-1} = \frac{1051{,}51}{199} = 5{,}284;$$

смещенная оценка среднеквадратического отклонения:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n}} = \sqrt{\frac{1051{,}51}{200}} = 2{,}293;$$

несмещенная оценка среднеквадратического отклонения:

$$S_x = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1051{,}51}{199}} = 2{,}299; \tag{72}$$

$$V = \frac{S_x}{\bar{x}} \cdot 100\% = \frac{2{,}299}{24{,}80} \cdot 100\% = 9{,}270\%;$$

$$A = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^3}{n \cdot S_x^3} = \frac{-1844{,}2}{200 \cdot 2{,}299^3} = -0{,}7589; \tag{73}$$

$$E = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4}{n \cdot S_x^4} - 3 = \frac{21\,612{,}6}{200 \cdot 2{,}299^4} = 3{,}8683 - 3 = 0{,}8683; \tag{74}$$

$$\mu_0 = 1;\ \mu_1 = 0;\ \mu_2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n} = \frac{1051{,}51}{200} = 5{,}258;$$

$$\mu_3 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^3}{n} = \frac{-1844{,}2}{200} = -9{,}221;$$

$$\mu_4 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4}{n} = \frac{21\,612{,}6}{200} = 108{,}1.$$

2.3. Структурные характеристики статистического ряда

Наряду со степенными средними и показателями вариации для характеристики экспериментальных данных используются так называемые структурные характеристики. Они обычно располагаются в определенных местах вариационного ряда. Ниже будут рассмотрены некоторые структурные характеристики вариационного ряда.

Мода. На рис. 3 изображен полигон распределения 218 сосновых стволов по диаметрам. Как видим, максимальное число наблюдений попало в интервал, имеющий значение 23,3 см (21,95–24,65). Этот класс называют модальным, а значение признака, которое наиболее часто встречается в выборке, – модой. Распределение, приведенное на рис. 3, называется *унимодальным*, так как оно имеет только один максимум – уже упоминавшийся класс со значением 23,3 см. Если полигон распределения имеет два или более максимума, то такое распределение называется *бимодальным* или *мультимодальным* соответственно.

В том случае, если анализируется дискретный признак и исходные данные сгруппированы в классы, созданные для каждого значения признака, мода непосредственно равна значению модального класса. Например, рассмотрим распределение сеголеток (первые две генерации) рыжей полевки по величине сентябрьских выводков (табл. 5). В данном примере модальный класс образуют 11 самок, имеющих по 5 эмбрионов. Мода для данного случая равна 5.

Если при анализе непрерывно варьирующего признака исходные данные сгруппированы в интервальный ряд, то мода может находиться в любом месте модального интервала. Ее местоположение можно оценить, смоделировав зависимость частоты от величины исследуемого признака в модальном и двух соседних с ним интервалах с помощью параболы второго порядка:

$$f = a_0 + a_1 \cdot x + a_2 \cdot x^2, \tag{75}$$

где a_0, a_1 и a_2 – постоянные коэффициенты; f – частота; x – исследуемый признак. Для того чтобы упростить дальнейшие выкладки, сделаем замену переменной:

$$z = \frac{x - x_m}{\lambda}; \qquad x = x_m + \lambda \cdot z, \tag{76}$$

где x_m – центр модального интервала, λ – величина интервала. С учетом (76) выражение (75) преобразуется к следующему виду:

$$\begin{aligned} f &= a_0 + a_1 \cdot (x_m + \lambda \cdot z) + a_2 \cdot (x_m + \lambda \cdot z)^2 = \\ &= a_0 + a_1 \cdot (x_m + \lambda \cdot z) + a_2 \cdot (x_m^2 + 2 \cdot x_m \cdot \lambda \cdot z + \lambda^2 \cdot z^2) = \\ &= a_0 + a_1 \cdot x_m + a_1 \cdot \lambda \cdot z + a_2 \cdot x_m^2 + 2 \cdot a_2 \cdot x_m \cdot \lambda \cdot z + a_2 \cdot \lambda^2 \cdot z^2 = \\ &= a_0 + a_1 \cdot x_m + a_2 \cdot x_m^2 + (a_1 \cdot \lambda + 2 \cdot a_2 \cdot x_m \cdot \lambda) \cdot z + a_2 \cdot \lambda^2 \cdot z^2. \end{aligned} \tag{77}$$

Обозначим:

$$\begin{aligned} b_0 &= a_0 + a_2 \cdot x_m^2; \\ b_1 &= a_1 \cdot \lambda + 2 \cdot a_2 \cdot x_m \cdot \lambda; \\ b_2 &= a_2 \cdot \lambda^2. \end{aligned} \tag{78}$$

Так как x_m и λ – константы, то b_0, b_1 и b_2 также являются константами. С учетом (78) выражение (77) можно переписать следующим образом:

$$f = b_0 + b_1 \cdot z + b_2 \cdot z^2. \tag{79}$$

Мода будет соответствовать локальному максимуму функции (79). Для того чтобы найти его местоположение, надо вычислить значение z_e, для которого производная функции (79) будет равна нулю. Найдем производную этой функции:

$$f' = b_1 + 2 \cdot b_2 \cdot z.$$

Затем, приравняв ее к нулю, найдем местоположение локального экстремума:

$$b_1 + 2 \cdot b_2 \cdot z_e = 0,$$

откуда

$$z_e = -\frac{b_1}{2 \cdot b_2}. \qquad (80)$$

Теперь нам надо оценить значение констант a_0, a_1 и a_2. Для этого воспользуемся значениями частот в модальном и соседних интервалах. Подставив эти значения в формулу (79), получим систему уравнений:

$$\begin{cases} f_{m-1} = b_0 + b_1 \cdot z_{m-1} + b_2 \cdot z_{m-1}^2; \\ f_m = b_0 + b_1 \cdot z_m + b_2 \cdot z_m^2; \\ f_{m+1} = b_0 + b_1 \cdot z_{m+1} + b_2 \cdot z_{m+1}^2, \end{cases} \qquad (81)$$

где f_{m-1} – частота класса, предшествующего модальному; f_m – частота модального интервала; f_{m+1} – частота класса, следующего за модальным; z_{m-1}, z_m и z_{m+1} – значения переменной z, соответствующие центрам предшествующего модальному, модального и следующего за модальным интервалов. С учетом (76) получим

$$z_{m-1} = \frac{x_{m-1} - x_m}{\lambda} = -1;$$

$$z_m = \frac{x_m - x_m}{\lambda} = 0; \qquad (82)$$

$$z_{m+1} = \frac{x_{m+1} - x_m}{\lambda} = 1.$$

Подставляя значения переменной z из (82) в систему уравнений (81), получим

$$\begin{cases} f_{m-1} = b_0 - b_1 + b_2; \\ f_m = b_0; \\ f_{m+1} = b_0 + b_1 + b_2. \end{cases} \qquad (83)$$

Вычитая второе уравнение системы (83) из первого и третьего и перенося f_m в правые части, получим новую систему из двух уравнений:

$$\begin{cases} f_{m-1} = f_m - b_1 + b_2; \\ f_{m+1} = f_m + b_1 + b_2. \end{cases} \qquad (84)$$

Далее найдем сумму и разность уравнений системы (84):

$$\begin{cases} f_{m+1} + f_{m-1} = 2 \cdot f_m + 2 \cdot b_2; \\ f_{m+1} - f_{m-1} = 2 \cdot b_1. \end{cases} \tag{85}$$

Преобразуем уравнения системы (85) следующим образом:

$$\begin{cases} 2 \cdot b_2 = f_{m+1} - 2 \cdot f_m + f_{m-1}; \\ b_1 = \dfrac{f_{m+1} - f_{m-1}}{2}. \end{cases} \tag{86}$$

Используя уравнения системы (86), мы можем преобразовать выражение (80) следующим образом:

$$z_e = -\frac{b_1}{2 \cdot b_2} = \frac{f_{m-1} - f_{m+1}}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m+1})}. \tag{87}$$

Таким образом, мы выразили значение переменной z, соответствующее локальному экстремуму, через значения частот в модальном и соседних интервалах. Теперь нам надо убедиться, что этот локальный экстремум является локальным максимумом. Для этого найдем вторую производную функции (79):

$$f'' = 2 \cdot b_2.$$

С учетом первого уравнения системы (86) получим

$$f'' = f_{m+1} - f_m + f_{m-1} - f_m.$$

Так как частоты соседних с модальным классов меньше, чем частота модального класса, то вторая производная функции (79) в точке z_e, как, впрочем, и в любой другой точке, будет отрицательна. Это говорит нам о том, что в точке z_e функция (79) имеет локальный максимум.

Теперь, используя (76), сделаем обратную подстановку и преобразуем выражение (87) к виду

$$z_e = \frac{x_e - x_m}{\lambda} = \frac{f_{m-1} - f_{m+1}}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m-1})},$$

откуда мода будет равна

$$Mo = x_e = x_m + \lambda \cdot \frac{f_{m-1} - f_{m+1}}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m-1})}. \tag{88}$$

Если использовать не центр модального интервала, а его нижнюю границу $x_{,}$, то выражение (88) можно преобразовать к виду

$$Mo = x_{\text{н}} + \frac{\lambda}{2} + \lambda \cdot \frac{f_{m-1} - f_{m+1}}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m-1})} =$$

$$= x_{\text{н}} + \lambda \cdot \frac{f_{m+1} - 2 \cdot f_m + f_{m-1} + f_{m-1} - f_{m+1}}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m-1})} =$$

$$= x_{\text{н}} + \lambda \cdot \frac{2 \cdot f_{m-1} - 2 \cdot f_m}{2 \cdot (f_{m+1} - 2 \cdot f_m + f_{m-1})} =$$

$$= x_{\text{н}} + \lambda \cdot \frac{f_m - f_{m-1}}{2 \cdot f_m + f_{m-1} - f_{m+1}}.$$

В качестве примера найдем моду для распределения деревьев по диаметру на высоте груди, приведенному в табл. 11. Модальным в этом распределении является интервал 21,95–24,65. Центр этого интервала находится в точке 23,3. Используя формулу (88), получим

$$Mo = 23{,}3 + 2{,}6 \cdot \frac{29 - 39}{2 \cdot (39 - 2 \cdot 40 + 29)} = 23{,}3 + 1{,}1 = 24{,}4.$$

Медиана. Положение экспериментальных данных достаточно хорошо характеризуется различными степенными средними. Однако в случае малой выборки на величину этих статистик могут оказывать довольно значительное влияние крайние варианты, которые являются наименее характерными элементами выборки. Этого недостатка лишена медиана, значение которой определяется наиболее типичными элементами выборки. Медиана – это значение признака, которое делит всю выборку на две равные части. Половина вариант имеют значения меньшие, чем медиана, а половина – большие.

Проще всего значение медианы определяется в случае несгруппированного набора данных. Для того чтобы определить медиану, надо предварительно упорядочить все элементы выборки по возрастанию (*ранжировать*). В том случае, если число элементов в выборке нечетное, мода будет равна варианте, имеющей в ранжированном ряду порядковый номер

$$j = \frac{n+1}{2}, \tag{89}$$

т. е.

$$Me = x_j,$$

где $j = \dfrac{n+1}{2}.$

Действительно, в силу того, что число элементов в выборке нечетное, величина j (89) будет целой, причем как слева от x_j, так и справа в ранжированном ряду будет находиться по $(n-1)/2$ элементов, т. е. x_j делит выборку на две группы с одинаковым числом наблюдений.

В том случае, если выборка будет иметь четное число наблюдений, медиана будет находиться посередине между $n/2$-м и $n/2+1$-м наблюдением, т. е.

$$Me = \frac{x_j + x_{j+1}}{2}, \tag{90}$$

где

$$j = \frac{n}{2}.$$

В качестве примера определим медиану для данных, характеризующих возобновление березы (табл. 3). Предварительно эту выборку необходимо упорядочить по числу порослевин (табл. 24). Данная выборка состоит из четного числа вариант (20), поэтому можно воспользоваться формулой (90). Половина числа наблюдений анализируемой выборки равна

$$j = \frac{20}{2} = 10.$$

Подставляя значения вариант $x_{10} = 7$ и $x_{11} = 8$ в формулу (90), получим

$$Me = \frac{7+8}{2} = 7{,}5.$$

Таблица 24. **Характеристика возобновления березы (ранжированный ряд)**

№ ствола в ранжированном ряду	№ ствола	Число порослевин	№ ствола в ранжированном ряду	№ ствола	Число порослевин
1	15	2	11	2	8
2	16	2	12	10	10
3	9	3	13	3	13
4	14	3	14	12	14
5	17	3	15	8	16

6	18	4	16	1	17
7	6	5	17	13	17
8	11	5	18	4	20
9	19	5	19	7	31
10	5	7	20	20	20

В том случае, если медиану надо определить для сгруппированного набора данных, начинают с того, что определяют, в каком классе она находится. Проще всего это сделать, если имеются в наличии накопленные частоты вариационного ряда. Класс, в котором находится медиана (медианный класс) – это первый класс, у которого накопленная частота окажется больше, чем $n/2$. В случае дискретной вариации, когда данные группировались в безинтервальный вариационный ряд, значение этого класса и будет медианой. Если группировка производилась в интервальный вариационный ряд, предполагая, что внутри медианного интервала наблюдения располагаются равномерно, медиану можно определить по формуле

$$Me = x_{\text{н}}^{Me} + \lambda \cdot \left(\frac{\frac{n}{2} - \sum_{i=1}^{j-1} f_i}{f_j} \right), \qquad (91)$$

где $x_{\text{н}}^{Me}$ – нижняя граница классового интервала; j – номер медианного интервала; $\sum_{i=1}^{j-1} f_i$ – накопленная частота предшествующего медианному класса; f_j – частота медианного класса.

Рассмотрим определение медианы в случае безинтервального вариационного ряда на примере распределения сеголеток рыжей полевки по величине сентябрьских выводков (табл. 5). Медианным в данном случае является класс, к которому относятся полевки, имеющие по 5 эмбрионов, так как это первый класс, чья накопленная частота (30) больше, чем половина объема выборки: $n/2 = 44/2 = 22$. Так как мы рассматриваем дискретный признак, сгруппированный в безинтервальный вариационный ряд, то значению этого медианного класса и будет соответствовать медиана, т. е.

$Me = 5.$

Теперь определим медиану для случая, когда признак сгруппирован в интервальный вариационный ряд. В табл. 11

приведено распределение деревьев по диаметру на высоте груди. Половина объема данной выборки составит

$$\frac{n}{2} = \frac{218}{2} = 109.$$

Первым интервалом, чья накопленная частота превышает это значение (медианным интервалом), будет интервал 21,95–24,65. Подставляя необходимые значения в формулу (91), получим

$$Me = x_{\text{н}}^{Me} + \lambda \cdot \frac{\frac{n}{2} - \sum_{i=1}^{j-1} f_i}{f_j} = 22{,}0 + 2{,}7 \cdot \frac{\frac{218}{2} - 77}{40} =$$

$$= 22{,}0 + 2{,}7 \cdot \frac{32}{40} = 24{,}16.$$

Квантили. Медиана делит вариационный ряд на две равные части. В более общем случае мы можем разделить вариационный ряд на две неравные части в любом соотношении. Статистики, которые отделяют от вариационного ряда определенную часть его членов, называются квантилями.

Квантили, которые отделяют от вариационного ряда 1, 2, ..., 99 % его членов, называются *перцентилями*. С помощью 99 перцентилей $P_1, P_2, ..., P_{99}$ вариационный ряд делится на 100 равных частей. Девять статистик, которые делят вариационный ряд на десять одинаковых частей, называются *децилями*. *Квартилями* называют три квантиля (Q_1, Q_2 и Q_3), которые делят вариационный ряд на четыре равные части. Они соответствуют перцентилям, отделяющим от ранжированного ряда наблюдений 25, 50 и 75% вариант соответственно:

$$Q_1 = P_{25};$$

$$Q_1 = P_{50};$$

$$Q_1 = P_{75}.$$

Кроме того, квартиль и перцентиль, делящие вариационный ряд на две равные части, соответствуют медиане ряда наблюдений:

$$Q_2 = P_{50} = Me.$$

На практике чаще всего используют перцентили P_3, P_{10}, P_{25}, P_{50}, P_{75}, P_{90} и P_{97}. Определяют квантили аналогично тому, как определяют медиану вариационного ряда. В том случае, если анализируется интервальный вариационный ряд, можно воспользоваться формулой

$$P_L = x_{\text{н}}^L + \lambda \cdot \left(\frac{\frac{L \cdot n}{100} - \sum_{i=1}^{j-1} f_i}{f_L} \right), \qquad (92)$$

где P_L – квантиль, отделяющий от ранжированного ряда L процентов наблюдений; $x_{\text{н}}^L$ – нижняя граница интервала, в который попадает квантиль P_L; j – номер интервала, в который попадает квантиль P_L; L – процент наблюдений в выборке, которые меньше, чем квантиль P_L; $\sum_{i=1}^{j-1} f_i$ – накопленная частота интервала, предшествующего интервалу, в котором находится квантиль P_L. Для того чтобы определить, в каком интервале находится квантиль, следует воспользоваться накопленными частотами ряда распределения. Первый интервал, у которого накопленная частота окажется больше, чем величина $L \cdot n/100$, и будет таким классом.

В качестве примера найдем квартили распределения деревьев по диаметру на высоте груди (табл. 11). Квартиль Q_1 попадает в интервал 19,25–21,95, так как это первый интервал, чья накопленная частота (77) оказалась больше величины:

$$\frac{L \cdot n}{100} = \frac{25 \cdot 218}{100} = 54{,}5.$$

Подставляя необходимые значения в формулу (92), получим

$$Q_1 = P_{25} = 19{,}3 + 2{,}7 \cdot \left(\frac{\frac{25 \cdot 218}{100} - 48}{29} \right) = 19{,}91.$$

Аналогичным образом определяем значения остальных двух квартилей распределения:

$$Q_2 = P_{50} = 22{,}0 + 2{,}7 \cdot \left(\frac{\frac{50 \cdot 218}{100} - 77}{40} \right) = 24{,}16 = Me$$

и

$$Q_3 = P_{75} = 27{,}4 + 2{,}7 \cdot \left(\frac{\dfrac{75 \cdot 218}{100} - 156}{28} \right) = 28{,}12.$$

3. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Лесная биометрия имеет дело с признаками, характеризующими различным образом лес в целом или отдельные его элементы. Для природных объектов свойственна изменчивость, т. е. признаки, которые их характеризуют, изменяются от объекта к объекту. Значение таких признаков определяется влиянием на объект множества различных, как правило, неизвестных случайных причин. В связи с этим признаки, характеризующие природные объекты, сами являются случайными величинами.

3.1. Функция распределения

Любая случайная величина подчинена какому-либо закону распределения, который может быть описан с помощью функции, определяемой соотношением

$$F_X(x) = P(X < x)$$

и называемой функцией распределения величины X.

Функция распределения обладает рядом свойств.

7. Функция распределения может принимать значения в интервале от 0 до 1 включительно:

$$0 \le F(x) \le 1.$$

8. Функция распределения является неубывающей функцией, т. е. для любых x_1 и x_2 таких, что $x_1 \le x_2$, справедливо неравенство

$$F(x_1) \le F(x_2),$$

так как вероятность того, что случайная величина попадет на отрезок $]-\infty; x_2[$ не меньше, чем вероятность попадания этой случайной величины на отрезок $]-\infty; x_1[$, если $x_2 \ge x_1$.

9. При x, стремящемся к $-\infty$, функция распределения стремится к нулю:

$$\lim_{x \to -\infty} F(x) = 0,$$

а при x, стремящемся к $+\infty$, функция распределения стремится к единице:

$$\lim_{x \to +\infty} F(x) = 1.$$

Кроме того, следует заметить, что величина $F(b)-F(a)$ представляет собой вероятность того, что случайная величина X примет значение, принадлежащее интервалу $a \le X < b$. Если верхняя граница b такого интервала равна $+\infty$, то событие, состоящее в том, что случайная величина попадает в интервал $[a;+\infty[$ есть не что иное, как событие, состоящее в том, что случайная величина окажется больше или равной величине a. Вероятность такого события равна

$$P(X \ge x) = F(+\infty) - F(x),$$

или с учетом свойства 3 функции распределения

$$P(X \ge x) = 1 - F(x).$$

3.2. Дискретные случайные величины

Случайные величины могут быть дискретными или непрерывными.

Дискретными называются случайные величины, принимающие конечное или счетное множество[2] значений. В качестве примера возьмем число порослевин от основания усохших после пожара берез или число яиц в кладках соснового пилильщика и т. д.

Известно много законов распределения дискретных случайных величин. В качестве примера рассмотрим биномиальное распределение.

3.2.1. Биномиальное распределение

Представим себе такую ситуацию: создаются лесные культуры дуба черешчатого строчно-луночным посевом желудей. Желуди высеваются в лунки, расположенные строчками через 0,7 м с расстоянием между строчками 1,2 м. В каждую лунку кладется по пять желудей. Часть из них взойдет, а другая часть не даст всходов. Предположим, что для высеваемой партии желудей была определена всхожесть семян. Всхожесть – это отношение числа семян, которые дали всходы, к общему количеству высеяных семян, выраженное в процентах. Обозначим символом p всхожесть, деленную на 100, т. е. p – это отношение числа семян, которые дали всходы, к общему количеству высеяных семян или, другими словами, p – это

---

[2] Множество называется счетным, если можно установить взаимно однозначное соответствие между ним и множеством натуральных чисел.

вероятность того, что желудь взойдет. Попробуем определить вероятность того, что в лунке взойдет 5 желудей, 4, 3, 2, 1 или не взойдет ни одного. Для удобства пронумеруем все желуди, высаженные в лунку, номерами с 1-го по 5-й. Теперь рассмотрим, какие могут быть варианты прорастания семян в лунке (табл. 25). Как видим, вероятность того, что в лунке не прорастет ни одного желудя, составляет q^5, где $q = 1 - p$ – вероятность того, что желудь не прорастет. Может произойти 5 различных событий, в результате которых в лунке будет только один всход. Эти события состоят в том, что в лунке прорастет 1, 2, 3, 4 или 5-й желудь, а остальные не взойдут. Вероятность каждого из этих событий, согласно правилу умножения вероятностей, равна $p \cdot q^4$. Таким образом, лунка, где есть только один росток, может встретиться с вероятностью $5 \cdot p \cdot q^4$. Аналогичным образом, пользуясь данными табл. 25, получаем, что вероятность встретить лунку с двумя всходами будет равна $10 \cdot p^2 \cdot q^3$, с тремя всходами – $10 \cdot p^3 \cdot q^2$, с четырьмя – $5 \cdot p^4 \cdot q$ и с пятью – p^5

Таблица. 25. **Возможные варианты прорастания желудей в лунке**

Число взошедших желудей	Номера взошедших желудей	Номера невзошедших желудей	Вероятность события
0	–	1,2,3,4,5	q^5
1	1	2,3,4,5	$p \cdot q^4$
1	2	1,3,4,5	$p \cdot q^4$
1	3	1,2,4,5	$p \cdot q^4$
1	4	1,2,3,5	$p \cdot q^4$
1	5	1,2,3,4	$p \cdot q^4$
2	1,2	3,4,5	$p^2 \cdot q^3$
2	1,3	2,4,5	$p^2 \cdot q^3$
2	1,4	2,3,5	$p^2 \cdot q^3$
2	1,5	2,3,4	$p^2 \cdot q^3$
2	2,3	1,4,5	$p^2 \cdot q^3$
2	2,4	1,3,5	$p^2 \cdot q^3$

Число взошедших желудей	Номера взошедших желудей	Номера невзошедших желудей	Вероятность события
2	2,5	1,3,4	$p^2 \cdot q^3$
2	3,4	1,2,5	$p^2 \cdot q^3$
2	3,5	1,2,4	$p^2 \cdot q^3$
2	4,5	1,2,3	$p^2 \cdot q^3$
3	1,2,3	4,5	$p^3 \cdot q^2$
3	1,2,4	3,5	$p^3 \cdot q^2$
3	1,2,5	3,4	$p^3 \cdot q^2$
3	1,3,4	2,5	$p^3 \cdot q^2$
3	1,3,5	2,4	$p^3 \cdot q^2$
3	1,4,5	2,3	$p^3 \cdot q^2$
3	2,3,4	1,5	$p^3 \cdot q^2$
3	2,3,5	1,4	$p^3 \cdot q^2$
3	2,4,5	1,3	$p^3 \cdot q^2$
3	3,4,5	1,2	$p^3 \cdot q^2$
4	1,2,3,4	5	$p^4 \cdot q$
4	1,2,3,5	4	$p^4 \cdot q$
4	1,2,4,5	3	$p^4 \cdot q$
4	1,3,4,5	2	$p^4 \cdot q$
4	2,3,4,5	1	$p^4 \cdot q$
5	1,2,3,4,5	–	p^5

Число событий, в которых имеется 0, 1, 2, 3, 4 или 5 взошедших желудей, есть не что иное, как биномиальные коэффициенты (число сочетаний из 5 по 0, 1, 2, 3, 4 или 5):

$$C_5^i = \frac{5!}{i!\cdot(5-i)!}, \qquad i = 0, 1, \ldots, 5.$$

Учитывая последнюю формулу, вероятность того, что в лунке окажется m всходов, будет равна

$C_5^m \cdot p^m \cdot q^{5-m}$.

Если бы в каждую лунку высевали не 5, а n желудей, то вероятность того, что в лунке окажется m всходов, составила бы величину

$$C_n^m \cdot p^m \cdot q^{n-m}. \qquad (93)$$

Это выражение называется *формулой Бернулли*. В общем случае рассматриваемую выше ситуацию можно представить как n-кратное повторение эксперимента. Причем в результате каждого случайного эксперимента событие A может произойти с вероятностью p или не произойти с вероятностью $q = 1 - p$. Таким образом, мы получаем случайную величину, представляющую собой число появлений события A в описываемой серии, состоящей из n независимых экспериментов. Эта случайная величина может принимать значения 0, 1, ..., n. Распределение, которому подчинена данная случайная величина, называется биномиальным, а вероятность того, что эта случайная величина будет равна m, определяется выражением (93).

3.3. Непрерывные случайные величины

Непрерывными называются случайные величины, которые могут принимать любые значения в каком-либо конечном или бесконечном интервале. Непрерывными случайными величинами являются такие параметры, как высота дерева, объем ствола и т. д.

3.3.1. Плотность распределения вероятностей

Плотностью распределения вероятностей $P(x)$ непрерывной случайной величины X называется предел, если он существует, отношения вероятности попадания случайной величины X на интервал $]x, x + \Delta x[$, примыкающий к точке x, к длине этого интервала, когда последняя стремится к нулю:

$$P(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x}.$$

Кривая, изображающая плотность распределения вероятностей (плотность вероятности) непрерывной случайной величины, называется *кривой распределения*.

Плотность вероятности обладает следующими свойствами.$P(x) \ge 0$ для любого x, принадлежащего интервалу $]-\infty;+\infty[$.

11. $\int\limits_{-\infty}^{+\infty} P(x)dx = 1$, так как вероятность того, что непрерывная случайная величина принимает значения из интервала]–∞;+∞[, т.е. вероятность достоверного события равна единице.

12. $P(x)$ – непрерывна или кусочно-непрерывна.

3.4. Нормальное распределение

В качестве примера для непрерывной случайной величины рассмотрим нормальное распределение. Это распределение имеет важное значение в биометрии. На практике очень часто изучаемые случайные величины следуют этому закону.Функция плотности вероятностей нормально распределенной случайной величины X имеет вид

$$P(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{(x-m)^2}{2 \cdot \sigma^2}},$$

где m – математическое ожидание, а σ – среднее квадратическое отклонение случайной величины X. Эта функция называется кривой Гаусса. Она симметрична относительно прямой $x = m$ и имеет колоколообразную форму.

При изменении величины m кривая будет смещаться влево или вправо, не меняя своей формы. Параметр m характеризует положение кривой на оси абсцисс, в связи с этим его иногда называют *параметром сдвига*.

Стандартное отклонение σ определяет форму кривой Гаусса. При малых значениях этого параметра кривая имеет узкую и высокую вершину, а при больших значениях выглядит плосковершинной (рис. 12). Этот параметр иногда называют *параметром масштаба*.

Кривая Гаусса имеет максимум (т. max) в точке $x = m$, равный

$$P(m) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}},$$

и две точки перегиба (т. перегиба) с абсциссами $m - \sigma$ и $m + \sigma$ и ординатой

$$P(m - \sigma) = P(m + \sigma) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi \cdot e}}$$

(рис. 13).

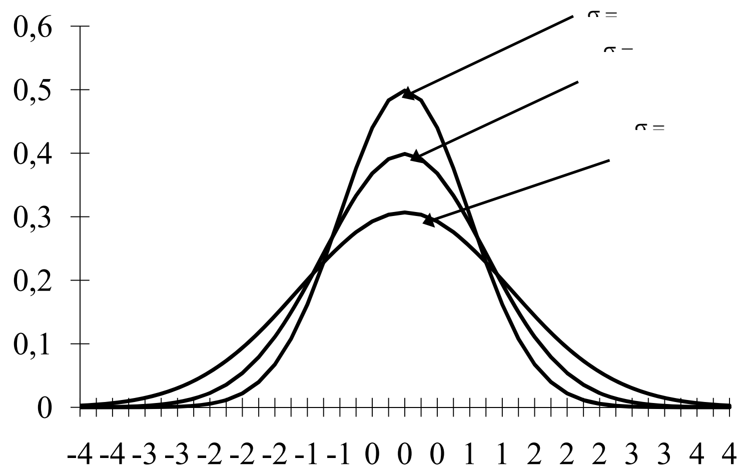


Рис 12. Нормальные кривые с разными значениями параметра формы

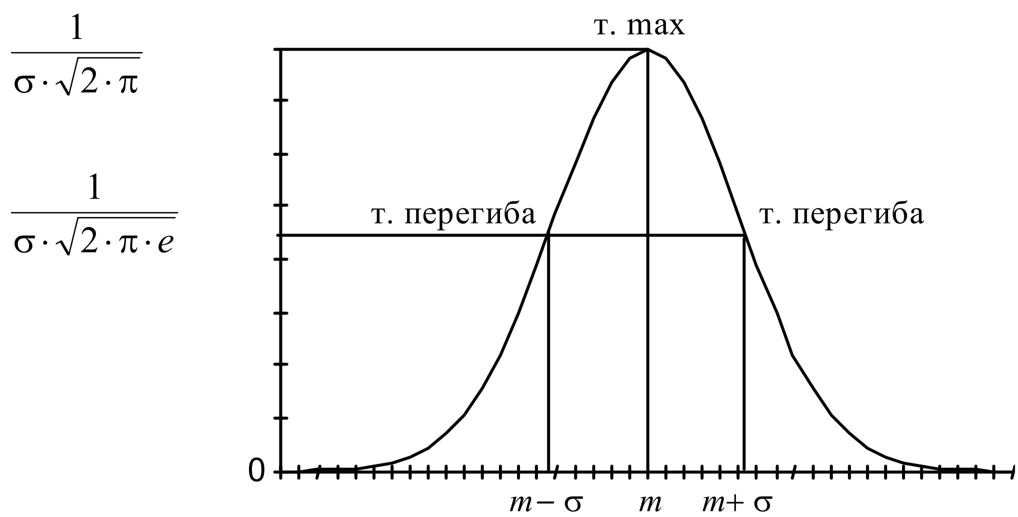


Рис. 13. Характерные точки кривой Гаусса

Площадь, ограниченная кривой Гаусса (как, впрочем, и любой другой кривой плотности распределения вероятностей) и осью абсцисс, равна 1. Причем основная площадь сосредоточена вблизи абсциссы, равной m. Так, например, площадь фигуры, ограниченной сверху кривой Гаусса, снизу – осью абсцисс, слева – вертикальной прямой, пересекающей ось абсцисс в точке $m - \sigma$, а справа – вертикальной прямой, пересекающей ось абсцисс в точке $m + \sigma$, равна 0,6827 (больше половины всей площади под кривой). Если слева и справа фигуру ограничить вертикальными прямыми, пересекающими ось абсцисс в точках $m - 2\sigma$ и $m + 2\sigma$, то ее площадь будет равна 0,9545, а если в точках $m - 3\sigma$ и $m + 3\sigma$, то 0,9973. Как известно, площадь криволинейной трапеции, ограниченной кривой плотности распределения вероятностей, осью абсцисс и прямыми $x = a$ и $x = b$, равна вероятности попадания непрерывной случайной величины в интервал $]a; b[$. Таким образом, случайная величина, подчиняющаяся закону нормального распределения с параметрами m и σ, с вероятностью 0,9973 попадет в интервал $]\, m - 3\sigma\, ;\, m + 3\sigma\, [$. Это значит, что вероятность ее отклонения от математического ожидания m на величину, превышающую три сигмы, очень мала. Иными словами, это означает, что такое событие практически невозможно. Это называется *правилом трех сигм.*

Функция нормального распределения $F(x)$ имеет вид

$$F(x) = \int_{-\infty}^{x} P(z)dz = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \int_{-\infty}^{x} e^{-\frac{(z-m)^2}{2\cdot\sigma^2}} dz\,.$$

Зная значения параметров m и σ, можно определить теоретические вероятности попадания исследуемой случайной величины в интервалы вариационного ряда (а следовательно, и его теоретические частоты), исходя из предположения, что она подчинена закону нормального распределения. Это позволит изобразить графически кривую нормального распределения и сравнить теоретические и эмпирические частоты вариационного ряда, на основании чего можно будет решить, следует эмпирическое распределение нормальному закону или нет.

Как отмечалось выше, вероятность попадания случайной величины в какой-нибудь интервал равна

$$P(a < X < b) = F(b) - F(a),$$

где $F(x)$ – функция распределения, a и b – границы интервала.

Для нормального распределения это можно записать следующим образом:

$$P(a < X < b) =$$

$$= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \int_{-\infty}^{b} e^{-\frac{(z-m)^2}{2\cdot\sigma^2}} dz - \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \int_{-\infty}^{a} e^{-\frac{(z-m)^2}{2\cdot\sigma^2}} dz. \qquad (94)$$

Интегралы, входящие в это выражение, нельзя выразить через элементарные функции, но их можно вычислить через специальную функцию

$$F(x)=\frac{1}{\sqrt{2\cdot\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt,$$

которая является интегральной функцией нормального распределения с параметрами $m = 0$ и $\sigma = 1$. Для этого следует перейти к нормированной случайной величине:

$$T=\frac{X-m}{\sigma}.$$

Преобразовав неравенство $a \le X < b$ соответствующим образом, получим

$$\frac{a-m}{\sigma}\le\frac{X-m}{\sigma}<\frac{b-m}{\sigma}.$$

Эти два неравенства равносильны, следовательно, их вероятности равны между собой:

$$P(a\le X<b)=P\left(\frac{a-m}{\sigma}\le\frac{X-m}{\sigma}<\frac{a-m}{\sigma}\right). \qquad (95)$$

Используя (95) и (94), получим

$$P(a\le X<b)=P\left(\frac{a-m}{\sigma}\le\frac{X-m}{\sigma}<\frac{a-m}{\sigma}\right)=P(t_1\le T<t_2)= \qquad (96)$$

$$=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{b} e^{-\frac{(z-m)^2}{2\cdot\sigma^2}}dz-\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{a} e^{-\frac{(z-m)^2}{2\cdot\sigma^2}}dz=$$

$$=\left|\begin{matrix}\dfrac{z-m}{\sigma}=t\\ dz=\sigma\cdot dt\end{matrix}\right|=\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\frac{b-m}{\sigma}} e^{-\frac{t^2}{2}}dz-\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\frac{a-m}{\sigma}} e^{-\frac{t^2}{2}}dz=$$

$$=\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{t_2} e^{-\frac{t^2}{2}}dz-\frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{t_1} e^{-\frac{t^2}{2}}dz=\Phi(t_2)-\Phi(t_1).$$

Используя (96) и статистические таблицы, мы можем вычислить теоретические частоты вариационного ряда, предполагая, что исследуемая случайная величина распределена по нормальному закону.

Выполним эту работу для вариационных рядов по диаметру и высоте, приведенных в табл. 12 и 13. Для вычислений составим вспомогательную таблицу (табл. 26). С учетом того, что оценкой параметров нормального распределения методом моментов являются среднеквадратическое отклонение и среднее арифметическое, вычислим нормированные нижнюю и верхнюю границы интервалов следующим образом:

$$t_i^{\text{н}} = \frac{x_i - \frac{\lambda}{2} - \bar{x}}{\sigma}; \; t_i^{\text{в}} = \frac{x_i + \frac{\lambda}{2} - \bar{x}}{\sigma}.$$

Таблица 26. **Вычисление теоретических частот для функции нормального распределения (диаметры)**

x_i	f_i	$t_i^{\text{н}}$	$t_i^{\text{в}}$	$\Phi(t_i^{\text{н}})$	$\Phi(t_i^{\text{в}})$	P_i	$\tilde{f}_i$	$\Delta = f_i - \tilde{f}_i$
14,45	0	–2,52	–2,09	0,000	0,018	0,018	3,6	–3,6
17,65	3	–2,09	–1,66	0,018	0,048	0,030	6,0	–3,0
20,85	11	–1,66	–1,23	0,048	0,106	0,058	11,6	–0,6
24,05	29	–1,23	–0,80	0,106	0,212	0,106	21,2	7,8
27,25	39	–0,80	–0,37	0,212	0,356	0,144	28,8	10,2
30,45	32	–0,37	0,06	0,356	0,524	0,168	33,6	–1,6
33,65	33	0,06	0,49	0,524	0,688	0,164	32,8	0,2
36,85	23	0,49	0,92	0,688	0,821	0,133	26,6	–3,6
40,05	10	0,92	1,35	0,821	0,911	0,090	18,0	–8,0
43,25	9	1,35	1,78	0,911	0,961	0,050	10,0	–1,0
46,45	3	1,78	2,21	0,961	0,986	0,025	5,0	–2,0
49,65	4	2,21	2,64	0,986	0,996	0,010	2,0	2,0
52,85	2	2,64	3,07	0,996	0,999	0,003	0,6	1,4
56,05	2	3,07	3,49	0,999	1,000	0,001	0,2	1,8
59,25	0	3,49	3,92	1,000	1,000	0,000	0,0	0,0
Сумма	200					1,000	200,0	0,0

В отличие от анализируемого вариационного ряда, нормальное распределение определено на интервале от –∞ до +∞. Для того чтобы области определения эмпирического и нормального распределения сделать одинаковыми, добавим дополнительные интервалы перед первым интервалом с границами от –∞ до нижней границы первого интервала и после последнего интервала с границами от верхней границы последнего интервала до +∞. Эмпирические частоты этих дополнительных интервалов будут равны нулю, так как в исходных данных нет ни одного наблюдения, которое было бы меньше нижней границы первого интервала или больше верхней границы последнего интервала. Значения функции нормированного нормального распределения для нижней $\Phi(t_i^{\text{н}})$ и верхней $\Phi(t_i^{\text{в}})$ границ интервалов можно найти с помощью табл. 2 приложения, используя в качестве аргументов значения $t_i^{\text{н}}$ и $t_i^{\text{в}}$ соответственно. В этой таблице значения функции распределения даны только для положительных аргументов. Если надо найти функцию распределения для отрицательного аргумента, следует воспользоваться соотношением Φ(–*x*) = 1 – Φ(*x*), которое справедливо, так как нормальное распределение является симметричным.

Вероятности для интервалов вариационного ряда легко вычислить как разность значений функции распределения для верхней и нижней границ:

$$P_i = \Phi(t_i^{\text{в}}) - \Phi(t_i^{\text{н}}).$$

Теперь можно найти теоретические частоты ряда:

$$f_i = n \cdot P_i.$$

Аналогичным образом можно вычислить теоретические частоты для вариационного ряда высот (табл. 27).

Таблица 27. **Вычисление теоретических частот для функции нормального распределения (высоты)**

x_i	f_i	$t_i^{\text{н}}$	$t_i^{\text{в}}$	$\Phi(t_i^{\text{н}})$	$\Phi(t_i^{\text{в}})$	P_i	$\widetilde{f}_i$	$\Delta = f_i - \widetilde{f}_i$
16,05	0	–4,02	–3,59	0	0	0	0	0
17,05	2	–3,59	–3,15	0	0,001	0,001	0,2	1,8
18,05	1	–3,15	–2,72	0,001	0,003	0,002	0,4	0,6
19,05	4	–2,72	–2,28	0,003	0,011	0,008	1,6	2,4
20,05	5	–2,28	–1,85	0,011	0,032	0,021	4,2	0,8
21,05	2	–1,85	–1,41	0,032	0,079	0,047	9,4	–7,4
22,05	13	–1,41	–0,98	0,079	0,164	0,085	17,0	–4
23,05	25	–0,98	–0,54	0,164	0,295	0,131	26,2	–1,2
24,05	31	–0,54	–0,11	0,295	0,456	0,161	32,2	–1,2
25,05	32	–0,11	0,33	0,456	0,629	0,173	34,6	–2,6
26,05	43	0,33	0,76	0,629	0,776	0,147	29,4	13,6
27,05	24	0,76	1,2	0,776	0,885	0,109	21,8	2,2
28,05	14	1,2	1,63	0,885	0,948	0,063	12,6	1,4
29,05	3	1,63	2,07	0,948	0,981	0,033	6,6	–3,6
30,05	1	2,07	2,5	0,981	0,994	0,013	2,6	–1,6
31,05	0	2,5	2,94	0,994	1	0,006	1,2	–1,2
Сумма	200					1,000	200,0	0,0

Последние колонки табл. 26 и 27, представляющие собой разность между эмпирическими и теоретическими частотами, дают нам информацию о близости теоретического (в данном случае нормального) и эмпирического распределений. Однако по данным отклонениям достаточно трудно принять решение о согласованности эмпирического и теоретического распределений. Более наглядную картину можно увидеть, изобразив эти распределения графически (рис. 14 и 15).

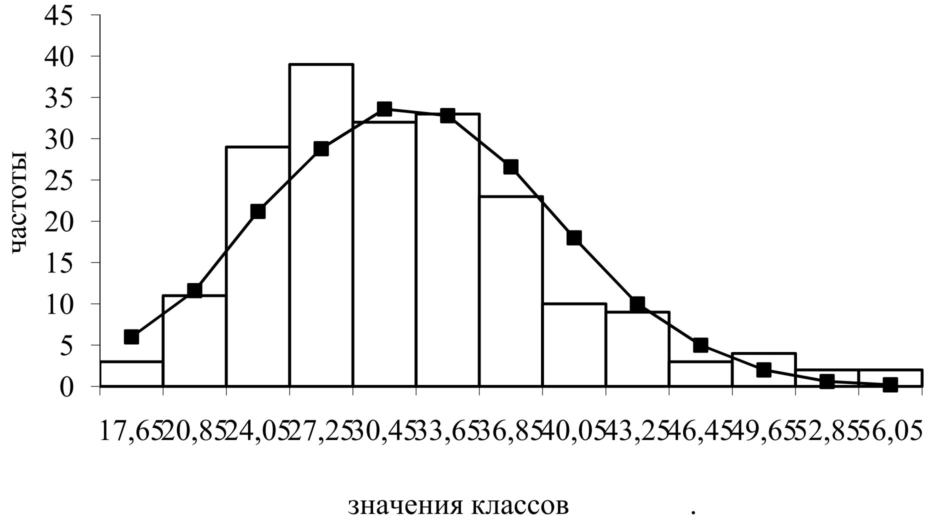


Рис. 14. Сравнение эмпирического и нормального распределений сосновых стволов по диаметрам

Рис. 15. Сравнение эмпирического и нормального распределений сосновых стволов по высотам

Однако такие сравнения распределений будут субъективными. Для того чтобы дать объективную оценку согласованности эмпирических и теоретических распределений, необходимо воспользоваться специальными методиками проверки статистических гипотез.

3.5. Некоторые распределения, используемые в лесном хозяйстве

Наряду с нормальным, в лесном деле используется еще целый ряд непрерывных распределений случайных величин. Рассмотрим некоторые из них.

Логнормальное распределение – это распределение случайной величины X, логарифм которой подчинен закону нормального распределения.

Область значений $0 \le x < +\infty$.

Функция плотности вероятности (рис. 16):

$$P(x) = \frac{1}{x \cdot \sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{-\left(\log\left(\frac{x}{m} \right) \right)^2}{2 \cdot \sigma^2} \right).$$

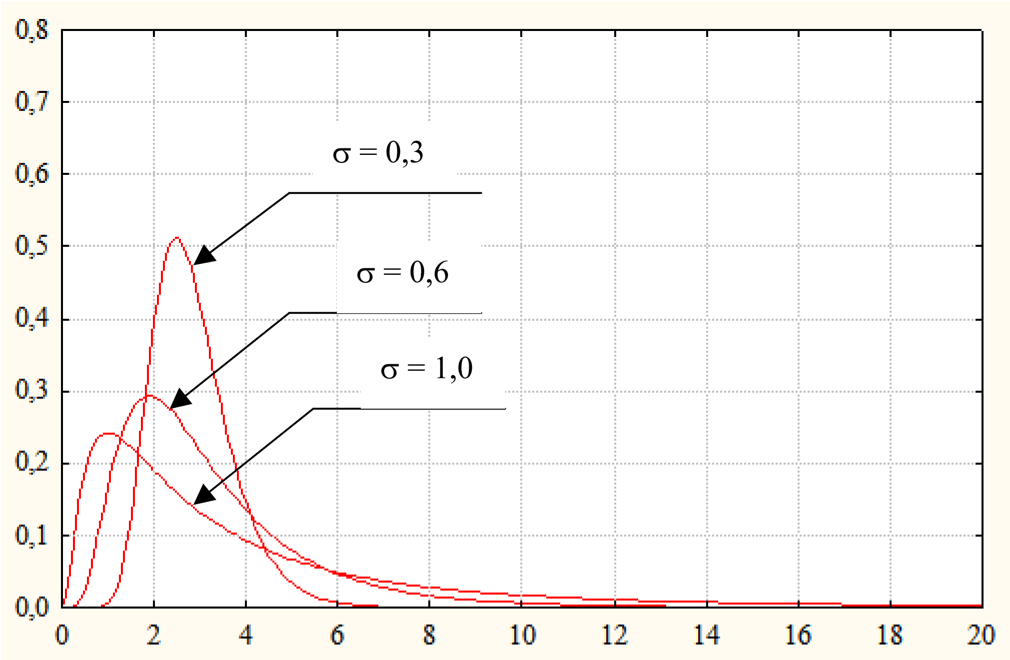


Рис. 16. Плотность распределения вероятностей (m = 1)

Функция распределения вероятностей для различных значений параметра σ и m = 1 приведена на рис. 17.

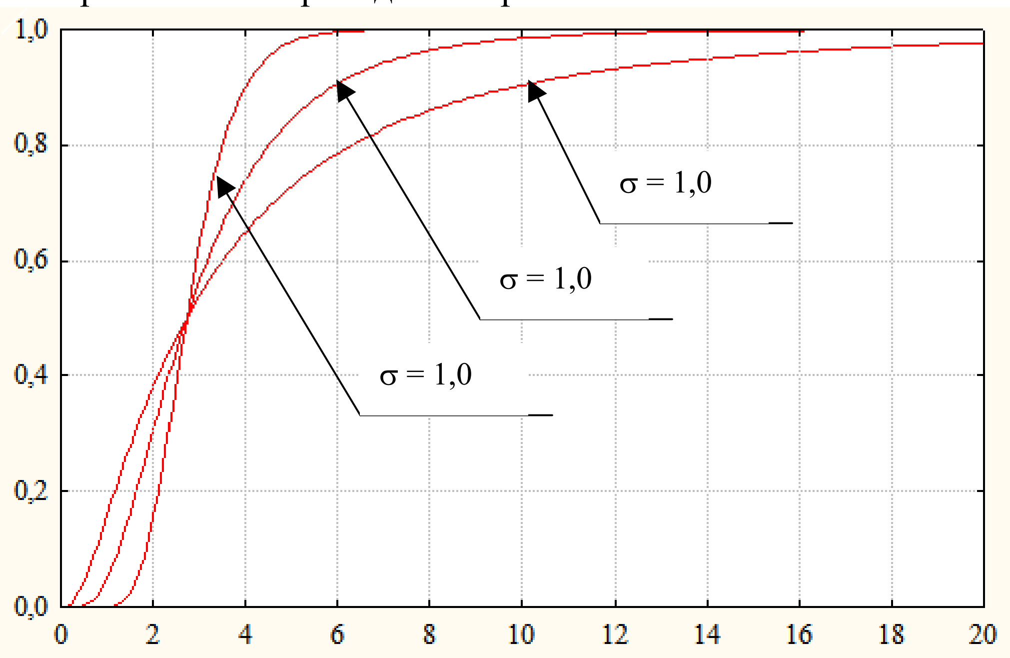


Рис. 17. Функция распределения вероятностей (m = 1)

Оценка параметров:

$\overline{m} = \exp(\overline{\mu})$;

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{n} (\log(x_i) - \overline{\mu})^2}{n-1},$$

где

$$\overline{\mu} = \frac{\sum_{i=1}^{n} \log(x_i)}{n}.$$

Гамма-распределение

Область значений $0 \le x < +\infty$.

Функция плотности вероятности (рис. 18):

$$P(x) = \left(\frac{x}{b}\right)^{c-1} \cdot \frac{\exp\left(-\frac{x}{b}\right)}{b \cdot \Gamma(c)},$$

где $\Gamma(c) = \int_0^{\infty} \exp(-u) \cdot u^{c-1} \cdot du$ – гамма-функция.

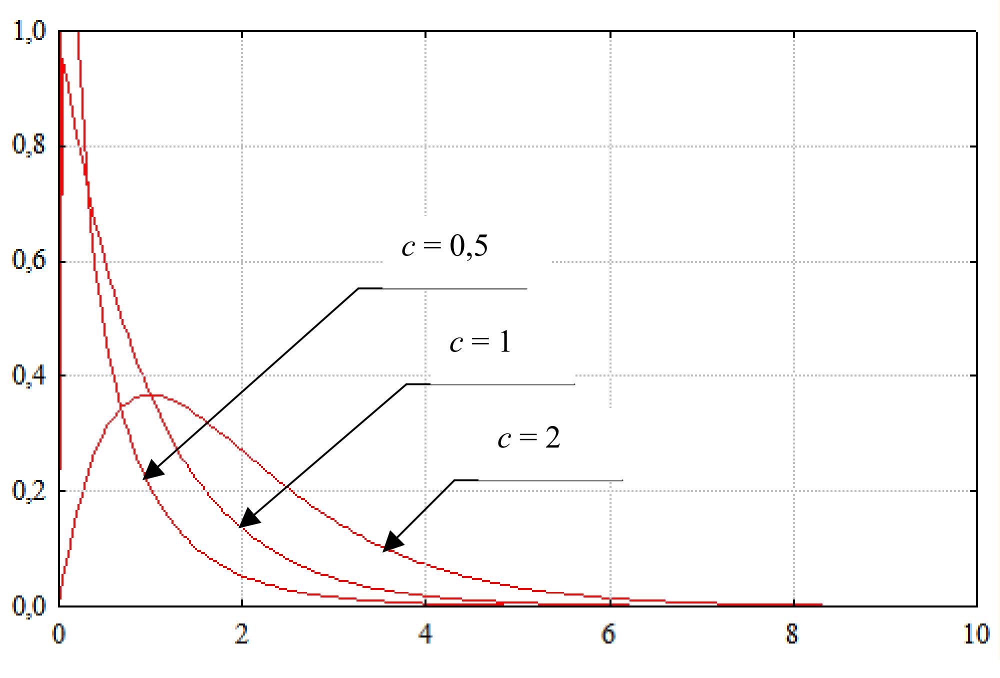


Рис. 18. Плотность распределения вероятностей (b = 1)

На рис. 19 приведена функция распределения вероятностей для различных значений параметра c и $b = 1$.

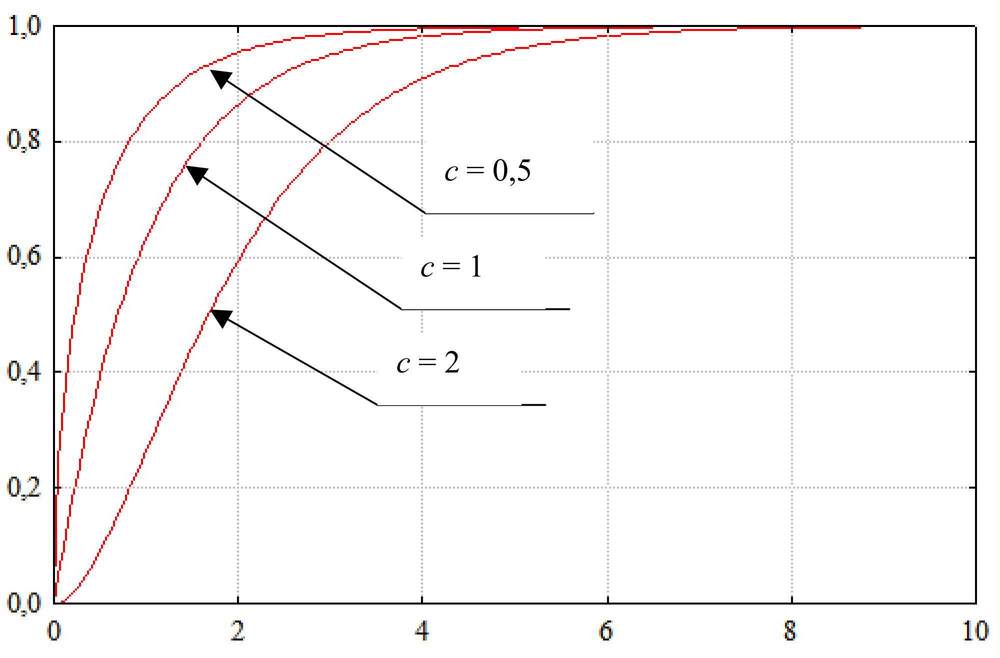


Рис. 19. Функция распределения вероятностей ($b = 1$)

Оценка параметров:

$$\bar{b} = \frac{S^2}{\bar{x}};$$

$$\bar{c} = \left(\frac{\bar{x}}{S}\right)^2,$$

где $\bar{x}$ – выборочное среднее; S^2 – выборочная дисперсия (без поправки).

Бета-распределение

Функция плотности вероятности (рис. 20, 22 и 24): $P(x) = \dfrac{x^{v-1} \cdot (1-x)^{w-1}}{\mathrm{B}(v, w)}$, $0 \le x \le 1$, $v > 0$, $w > 0$,где $\mathrm{B}(v,w)$ – бета-функция с параметрами v, w, задаваемая формулой:

$$\mathrm{B}(v,w) = \int_0^1 u^{v-1} \cdot (1-u)^{w-1} \cdot du.$$

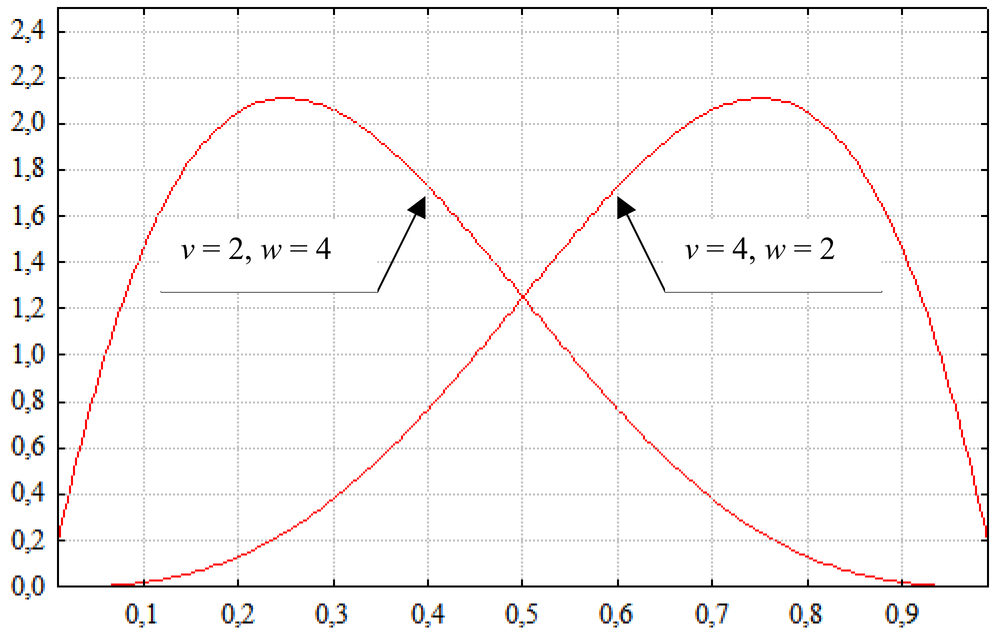


Рис. 20. Плотность распределения вероятностей

Вид функций распределения вероятностей для различных значений параметров v и w приведена на рис. 21, 23 и 25.

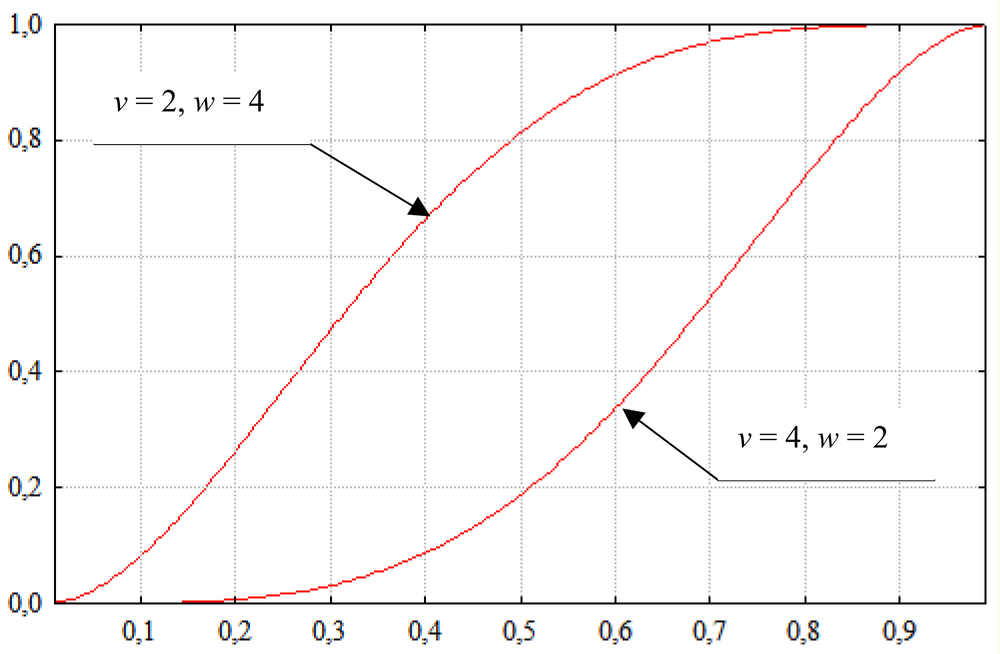


Рис. 21. Функция распределения вероятностей

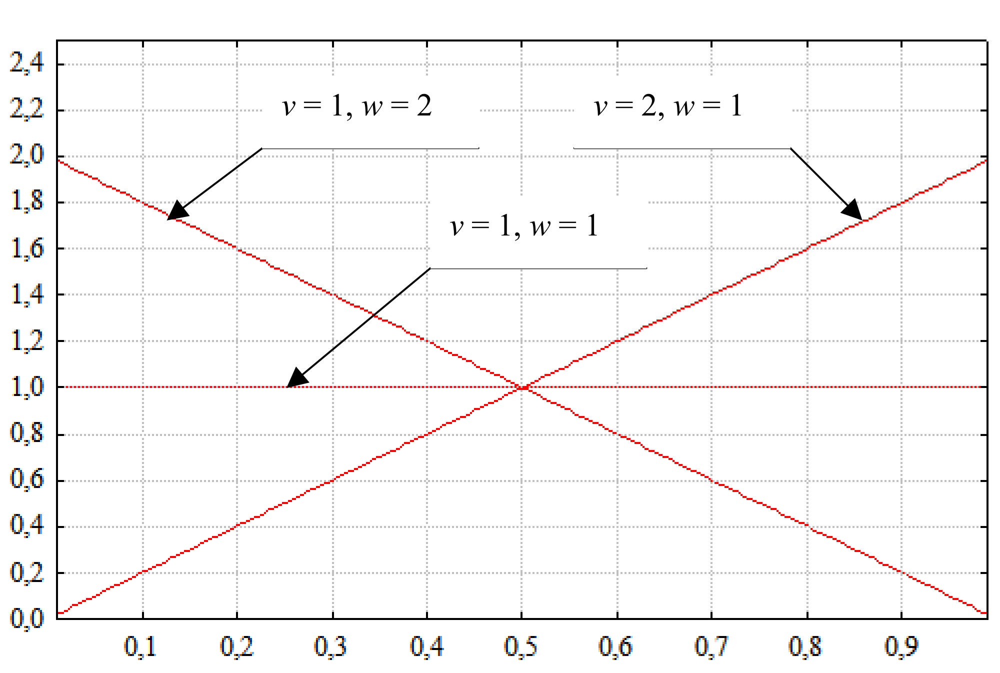


Рис. 22. Плотность распределения вероятностей

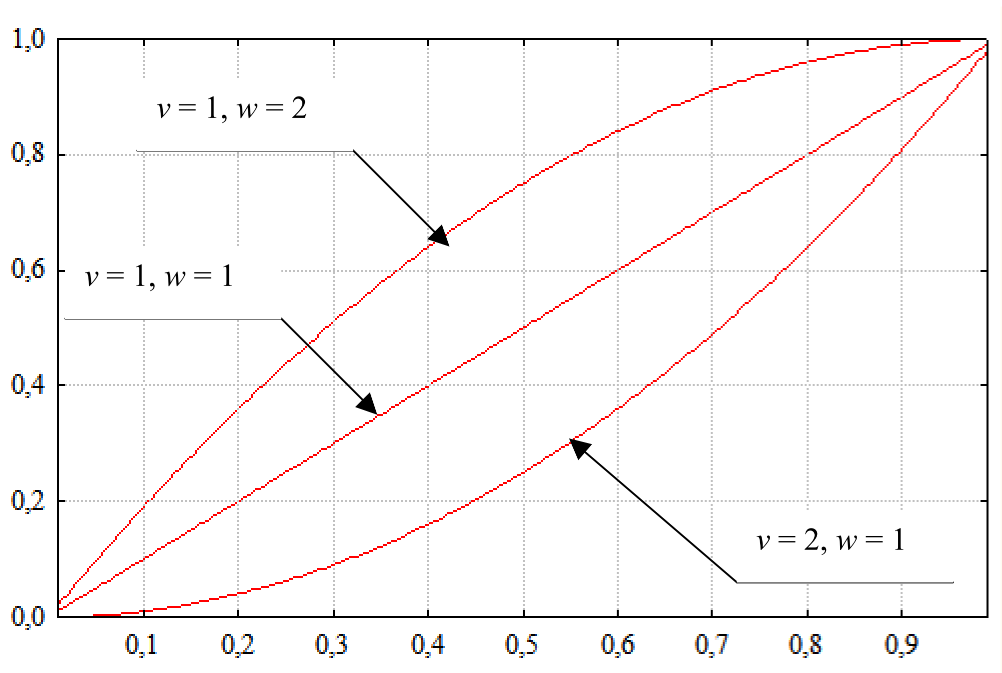


Рис. 23. Функция распределения вероятностей

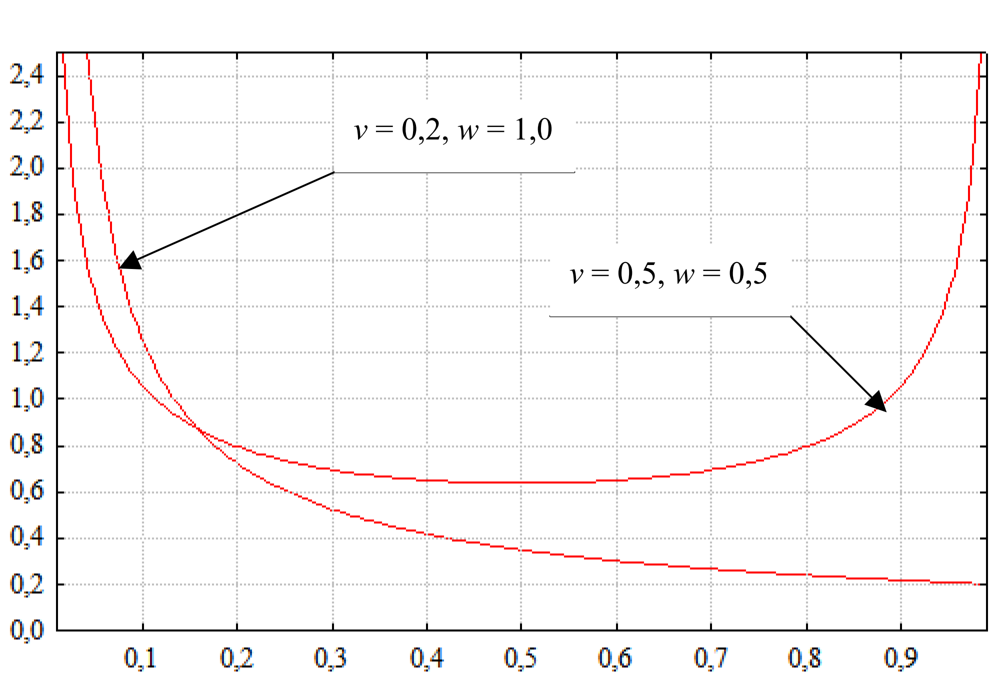


Рис. 24. Плотность распределения вероятностей

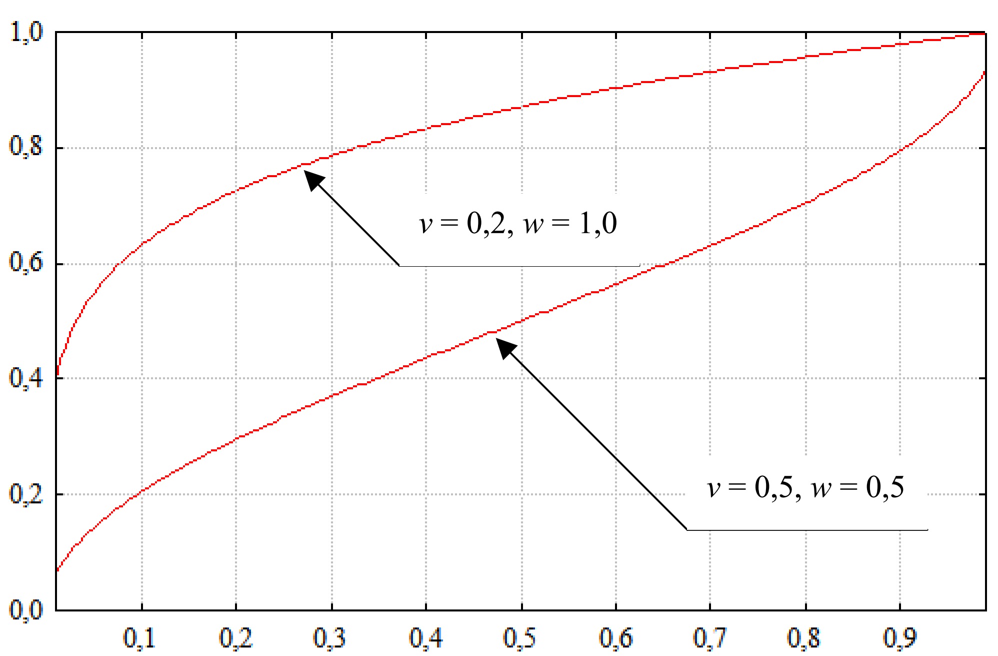


Рис. 25. Функция распределения вероятностей

Оценка параметров: $\overline{v} = \overline{x} \cdot \left(\dfrac{\overline{x} \cdot (1 - \overline{x})}{S^2} - 1 \right)$;

$$\overline{w} = (1-\overline{x}) \cdot \left(\frac{\overline{x} \cdot (1-\overline{x})}{S^2} - 1 \right),$$

где $\overline{x}$ – выборочное среднее; S^2 – выборочная дисперсия (без поправки Бесселя).

Распределение Грама-Шарлье (другие названия: обобщенное нормальное распределение или распределение типа A) представляет собой бесконечный ряд, использующий функцию нормального распределения и ее производные. Для практических целей обычно ограничиваются первыми тремя членами разложения.

Функция плотности вероятности:

$$P(x) = \varphi(x) - \frac{A}{6} \cdot \varphi^{\mathrm{III}}(x) + \frac{E}{24} \cdot \varphi^{\mathrm{IV}}(x),$$

где $\varphi(x)$ – функция плотности нормального распределения; $\varphi^{\mathrm{III}}(x)$, $\varphi^{\mathrm{IV}}(x)$ – третья и четвертая производные функции нормального распределения; A – коэффициент асимметрии; E – коэффициент эксцесса.При вычислении функции плотности вероятностей значения $\varphi(x)$, $\varphi^{\mathrm{III}}(x)$ и $\varphi^{\mathrm{IV}}(x)$ можно найти в статистических таблицах. При этом следует помнить следующие соотношения:

$\varphi(-x) = \varphi(x)$;

$\varphi^{\mathrm{III}}(-x) = -\varphi^{\mathrm{III}}(x)$;

$\varphi^{\mathrm{IV}}(-x) = \varphi^{\mathrm{IV}}(x)$.

Приблизительно вычислить теоретические частоты для распределения Гамма-Шарлье можно с помощью формулы

$$\tilde{f}_i = n \cdot \frac{\eta}{S_x} \cdot P(x), \tag{97}$$

где n – объем выборки; η – величина интервала; S_x – среднеквадратическое отклонение.

Однако следует помнить, что в статистических таблицах значения приведены для функции нормального распределения с параметрами $m = 0$ и $\sigma = 1$. Для того чтобы можно было воспользоваться этими таблицами, следует перейти к нормированной случайной величине:

$$T = \frac{X - m}{\sigma}.$$

Рассмотрим пример вычисления теоретических частот обобщенного нормального распределения на примере рядов распределения диаметров и высот деревьев в древостое (табл. 12 и 13). Для выполнения вычислений составим вспомогательную таблицу (табл. 28). С учетом того, что оценкой параметров нормального распределения методом моментов являются среднеквадратическое отклонение и среднее арифметическое, вычислим нормированные значения интервалов следующим образом:

$$t_i = \frac{x_i - \bar{x}}{S_x}.$$

После того как нормированные значения классов t_i вычислены, с помощью табл. 9 приложений следует найти для каждого интервала значения функции плотности нормального распределения, а также ее третьей и четвертой производных. Значения коэффициентов асимметрии и эксцесса возьмем из 2-й главы A = 0,8218 (70) и E = 0,7177 (71). Плотность распределения вероятностей для каждого интервала вычислим по формуле

$$P(t_i) = f(t_i) - \frac{A}{6} \cdot f^{\mathrm{III}}(t_i) + \frac{E}{24} \cdot f^{\mathrm{IV}}(t_i).$$

Таблица 28. **Вычисление теоретических частот для функции обобщенного нормального распределения (диаметры)**

x_i	f_i	t_i	$f(t_i)$	$f^{\mathrm{III}}(t_i)$	$f^{\mathrm{IV}}(t_i)$	$-\dfrac{A}{6}\cdot f^{\mathrm{III}}(t_i)$	$\dfrac{E}{24}\cdot f^{\mathrm{IV}}(t_i)$	$P(t_i)$	$\widetilde{f}_i$	$\Delta = f_i - \widetilde{f}_i$
17,65	3	–1,87	0,06 943	0,06 452	–0,39 946	–0,00 884	–0,01 195	0,04 864	4,2	–1,2
20,85	11	–1,44	0,14 146	–0,18 871	–0,72 736	0,02 585	–0,02 175	0,14 556	12,5	–1,5
24,05	29	–1,01	0,23 955	–0,47 903	–0,49 827	0,06 561	–0,01 490	0,29 026	24,9	4,1
27,25	39	–0,58	0,33 718	–0,52 091	0,36 913	0,07 135	0,01 104	0,41 957	36,0	3,0
30,45	32	–0,15	0,39 448	–0,17 618	0,13 038	0,02 413	0,00 390	0,42 251	36,3	–4,3
33,65	33	0,27	0,38 466	0,30 401	0,98 777	–0,04 164	0,02 954	0,37 256	32,0	1,0
36,85	23	0,70	0,31 225	0,54 863	0,09 371	–0,07 514	0,00 280	0,23 991	20,6	2,4
40,05	10	1,13	0,21 069	0,41 023	–0,63 857	–0,05 619	–0,01 910	0,13 540	11,6	–1,6
43,25	9	1,56	0,11 816	0,10 440	–0,67 104	–0,01 430	–0,02 007	0,08 379	7,2	1,8
46,45	3	1,99	0,05 508	–0,10 523	–0,27 970	0,01 441	–0,00 836	0,06 113	5,2	–2,2
49,65	4	2,42	0,02 134	–0,14 752	0,04 608	0,02 021	0,00 138	0,04 293	3,7	0,3
52,85	2	2,85	0,00 687	–0,10 034	0,13 910	0,01 374	0,00 416	0,02 477	2,1	–0,1
56,05	2	3,28	0,00 184	–0,04 682	0,09 970	0,00 641	0,00 298	0,01 123	1,0	1,0

Для вычисления теоретических частот воспользуемся формулой (97). В последнюю колонку табл. 28 запишем отклонения эмпирических частот ряда распределения диаметров деревьев от теоретических.

Аналогичным образом можно вычислить теоретические частоты для вариационного ряда высот (табл. 29).

Необходимые для вычисления значения коэффициентов асимметрии и эксцесса возьмем из главы 2 $A = –0{,}7589$ (73) и $E = 0{,}8683$ (74).

Последние колонки табл. 28 и 29, представляющие собой разность между эмпирическими и теоретическими частотами, дают нам информацию о близости обобщенного нормального и эмпирического распределений. Сравним эти отклонения с отклонениями, полученными для нормального распределения. Разница будет лучше видна, если мы изобразим эти распределения графически (рис. 26 и 27). Как видим, распределение диаметров деревьев в древостое имеет некоторую положительную асимметрию, а распределение высот – отрицательную. Кривая нормального распределения симметрична относительно среднего арифметического и, следовательно, она не может в полной мере отобразить все особенности распределения диаметров и высот деревьев в древостое. Обобщенное нормальное распределение является более гибким. Оно с успехом описывает распределения, имеющие как правостороннюю асимметрию (рис. 26, диаметры), так и левостороннюю (рис. 27, высоты).

Таблица 29. **Вычисление теоретических частот для функции обобщенного нормального распределения (высоты)**

x_i	f_i	t_i	$f(t_i)$	$f^{\mathrm{III}}(t_i)$	$f^{\mathrm{IV}}(t_i)$	$-\frac{A}{6}\cdot f^{\mathrm{III}}(t_i)$	$\frac{E}{24}\cdot f^{\mathrm{IV}}(t_i)$	$P(t_i)$	$\widetilde{f}_i$	$\Delta = f_i - \widetilde{f}_i$
17,05	2	–3,37	0,00 136	0,03 841	0,08 707	0,00 486	0,00 315	0,00 937	0,8	1,2
18,05	1	–2,94	0,00 530	0,08 788	0,13 691	0,01 112	0,00 495	0,02 137	1,9	–0,9
19,05	4	–2,50	0,01 753	0,14 242	0,07 997	0,01 801	0,00 289	0,03 843	3,3	0,7
20,05	5	–2,07	0,04 682	0,12 454	–0,20 363	0,01 575	–0,00 737	0,05 520	4,8	0,2
21,05	2	–1,63	0,10 567	–0,05 910	–0,62 161	–0,00 748	–0,02 249	0,07 570	6,6	–4,6
22,05	13	–1,20	0,19 419	–0,36 352	–0,69 255	–0,04 598	–0,02 506	0,12 315	10,7	2,3
23,05	25	–0,76	0,29 887	–0,55 023	–0,03 944	–0,06 959	–0,00 143	0,22 785	19,8	5,2
24,05	31	–0,33	0,37 780	–0,36 045	0,89 103	–0,04 559	0,03 224	0,36 445	31,7	–0,7
25,05	32	0,11	0,39 654	0,13 033	0,16 088	0,01 648	0,00 582	0,41 884	36,4	–4,4
26,05	43	0,54	0,34 482	0,50 431	0,46 048	0,06 379	0,01 666	0,42 527	37,0	6,0
27,05	24	0,98	0,24 681	0,49 332	–0,45 414	0,06 240	–0,01 643	0,29 278	25,5	–1,5
28,05	14	1,41	0,14 764	0,21 065	–0,73 466	0,02 664	–0,02 658	0,14 770	12,8	1,2
29,05	3	1,85	0,07 206	–0,05 633	–0,41 953	–0,00 712	–0,01 518	0,04 976	4,3	–1,3
30,05	1	2,28	0,02 965	–0,14 864	–0,03 461	–0,01 880	–0,00 125	0,00 960	0,8	0,2

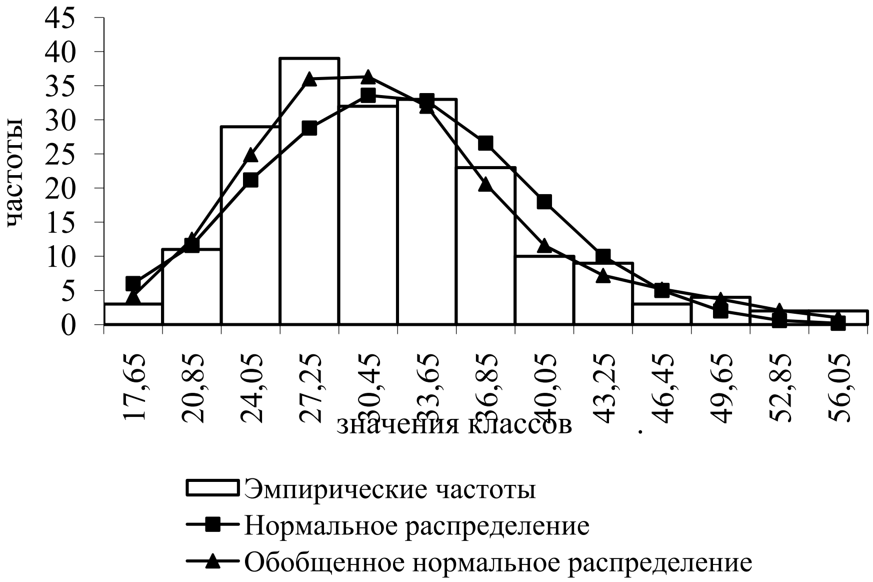


Рис. 26. Сравнение эмпирического нормального и обобщенного нормального распределений сосновых стволов по диаметрам

Рис. 27. Сравнение эмпирического нормального и обобщенного нормального распределений сосновых стволов по высотам

Анализ полученных результатов с помощью графиков весьма субъективен. Для того чтобы дать объективную оценку согласованности эмпирических и теоретических распределений, необходимо воспользоваться специальными методиками проверки статистических гипотез, которые будут рассматриваться в главе 5.

3.6. Система кривых Пирсона

Распределениями Пирсона называются непрерывные распределения, плотности вероятности которых являются решениями дифференциального уравнения

$$\frac{df(x)}{dx} = \frac{a_0 + a_1 \cdot x}{b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2} \cdot f(x).$$

Распределения Пирсона полностью определяются первыми четырьмя моментами. Если $a_1 = 1$, то: $a_0 = \frac{\mu_3 \cdot \left(\mu_4 + 3 \cdot \mu_2^2\right)}{A}$;

$$b_0 = -\frac{\mu_2 \cdot \left(4 \cdot \mu_2 \cdot \mu_4 - 3 \cdot \mu_3^2\right)}{A};$$

$$b_1 = -\frac{\mu_3 \cdot \left(\mu_4 + 3 \cdot \mu_2^2\right)}{2 \cdot A};$$

$$b_2 = -\frac{2 \cdot \mu_2 \cdot \mu_4 - 3 \cdot \mu_3^2 - 6 \cdot \mu_2^3}{A},$$

где $A = 10 \cdot \mu_4 \cdot \mu_2 - 18 \cdot \mu_2^3 - 12 \cdot \mu_3^2$; m_i – i-тый центральный момент.

В соответствии с распределением корней квадратного трехчлена $b_0 + b_1x + b_2x^2$ различают 12 типов распределений Пирсона.

Введем следующие обозначения:

$D = b_0 \cdot b_2 - b_1^2$ – дискриминант квадратного трехчлена;

$$\lambda = \frac{b_1^2}{b_0 \cdot b_2}.$$

Тип 1

$D < 0$, $\lambda < 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x_2 = (\alpha - x)\cdot(\beta - x)\cdot b_2$, $(\alpha, \beta > 0)$.

Функция плотности вероятности:

$$P(x) = \begin{cases} \dfrac{\alpha^{2\cdot m} \cdot \beta^{2\cdot n}}{(\alpha+\beta)^{m+n+1} \cdot \mathrm{B}(m+1, n+1)} \cdot (\alpha + x)^m \cdot (\beta - x)^n, \; x \in [-\alpha, \beta], \\ 0, \; x \notin [-\alpha, \beta], \end{cases}$$

где

$$\mathrm{B}(m+1, n+1) = \frac{\Gamma(m+1) \cdot \Gamma(n+1)}{\Gamma(m+n+2)}, \; m > -1, \; n > -1.$$

Распределениями Пирсона типа 1 являются бета-распределения.

Тип 2

Частный случай распределений типа 1. $\alpha = \beta$ (следовательно, $\lambda = 0$) и $m = n$. Кривая распределения симметрична относительно вертикальной оси.

Тип 3

$D < 0$, $\lambda = \infty$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x_2 = 2 \cdot (\alpha + x) \cdot b_1$.

Функция плотности вероятности:

$$P(x) = \begin{cases} \dfrac{k^{m+1}}{\Gamma(m+1)} \cdot (x+\alpha)^m \cdot e^{-k\cdot(\alpha+x)}, x > -\alpha, k > 0. \\ 0, x \le -\alpha. \end{cases}$$

Распределениями Пирсона типа 3 являются гамма-распределения.

Тип 4

$D > 0$, $0 < \lambda < 1$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x_2 = (x^2 + \alpha^2) \cdot b^2$.

Функция плотности вероятности:

$$P(x) = c \cdot \left(\alpha^2 + x^2\right)^{-m} \cdot \exp\left(-\nu \cdot \operatorname{arctg}\left(\frac{x}{\alpha}\right)\right),\ x \in (-\infty, \infty), m \ge 1/2,$$

где

$$c^{-1} = \int_{-\infty}^{\infty} \left(\alpha^2 + x^2\right)^{-m} \cdot \exp\left(-\nu \cdot \operatorname{arctg}\left(\frac{x}{\alpha}\right)\right) \cdot dx.$$

Тип 5

$D = 0$, $\lambda = 1$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = (x + \alpha)^2 \cdot b_2$.

Функция плотности вероятности:

$$P(x) = \begin{cases} \dfrac{\nu^{m-1}}{\Gamma(m-1)} \cdot x^{-m} \cdot e^{\frac{-\nu}{x}}, \nu > 0, m > 1, x > 0. \\ 0, x \le 0. \end{cases}$$

Тип 6

$D < 0$, $\lambda > 1$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = (x + \alpha) \cdot (x - \beta) \cdot b_2$.

Функция плотности вероятности:

$$P(x)=\begin{cases}\dfrac{(\alpha+\beta)^{-(m+n+1)}}{\mathrm{B}(-m-n-1,n+1)}\cdot(x+\alpha)^m\cdot(x-\beta)^n,\\ x>\beta, m-1>0, (\text{или} \quad m<-1), n>-1,\\ 0, x\le\beta,\end{cases}$$

Тип 7

$D > 0$, $\lambda = 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = (\alpha^2 + x^2)\cdot b_2$.

Функция плотности вероятности:

$$P(x)=\frac{\alpha}{\mathrm{B}(m-1/2,1/2)}\cdot(\alpha^2+x^2)^{-m},\ x\in(-\infty,\infty),\ m\ge 1/2.$$

Распределением Пирсона типа 7 является распределение Стьюдента.

Тип 8

$D < 0$, $\lambda < 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = x\cdot(\alpha + x)\cdot b_2$.

Функция плотности вероятности:

$$P(x)=\begin{cases}\dfrac{m+1}{\alpha^{m+1}}\cdot(x+\alpha)^m, x\in[-\alpha, 0], -1<m<0,\\ 0, x\notin[-\alpha, 0].\end{cases}$$

Тип 9

$D < 0$, $\lambda < 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = x\cdot(\alpha + x)\cdot b_2$.

Функция плотности вероятности:

$$P(x)=\begin{cases}\dfrac{m+1}{\alpha^{m+1}}\cdot(x+\alpha)^m, x\in[-\alpha, 0], m<-1,\\ 0, x\notin[-\alpha, 0].\end{cases}$$

Тип 10

$D = 0$, $\lambda = 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = b_0$, $a_1 = 0$.

Функция плотности вероятности:

$$P(x)=\begin{cases}\nu\cdot e^{-\nu\cdot x}, x>0, \nu>0,\\ 0, x\le 0.\end{cases}$$

Распределением Пирсона типа 10 является показательное распределение.

Тип 11

$D = 0$, λ неопределенно, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = b_0$, $a_1 \ne 0$.

Функция плотности вероятности:

$$P(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{x^2}{2 \cdot \sigma^2}}, \quad x \in (-\infty, \infty).$$

Распределением Пирсона типа 11 является нормальное распределение.

Тип 12

$D < 0$, $\lambda < 0$, $b_0 + 2 \cdot b_1 \cdot x + b_2 \cdot x^2 = (\alpha + x) \cdot (-\beta + x) \cdot b_2$, $(\alpha, \beta > 0)$.

Функция плотности вероятности:

$$P(x) = \begin{cases} \dfrac{\alpha^{-2 \cdot n} \cdot \beta^{2 \cdot n}}{(\alpha + \beta) \cdot \mathrm{B}(1 - n, n + 1)} \cdot (\alpha + x)^{-n} \cdot (\beta - x)^{n}, x \in [-\alpha, \beta], \\ 0, x \notin [-\alpha, \beta]. \end{cases}$$

4. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

Как известно, выборка x_1, x_2,..., x_n является реализацией случайного вектора (X_1, X_2,..., X_n). Это значит, что каждая числовая характеристика выборки есть реализация случайной величины, которая от выборки к выборке может принимать различные значения и, следовательно, сама является случайной. Такую случайную величину называют выборочной функцией, или статистикой, и обозначают $\tilde{a} = \tilde{a}(X_1, X_2, \ldots, X_n)$. Например, выборочными функциями являются среднее арифметическое $\bar{x}$, статистическая дисперсия S_x^2, мода *Мо*, медиана *Ме* и т. д.

Одна из главных задач анализа массовых данных заключается в том, чтобы на основании выборки сделать некоторые выводы о генеральной совокупности. Пусть требуется подобрать распределение для исследуемой случайной величины X по выборке x_1, x_2, ..., x_n, извлеченной из генеральной совокупности с неизвестной функцией распределения $F(x)$. Выбрав распределение (биномиальное, нормальное, показательное или др.) исходя из анализа выборки (например, по виду гистограммы или по виду полигона относительных частот), мы по данным выборки должны оценить параметры соответствующего распределения. Например, для нормального распределения нужно оценить параметры m и σ; для распределения Пуассона – параметры λ и т. д.

Существуют два основных метода получения оценок параметров генеральной совокупности по материалам выборки:

– точечное оценивание;
– интервальное оценивание.

Рассмотрим эти методы.

4.1. Точечное оценивание

Выборочная числовая характеристика, применяемая для получения оценки неизвестного параметра генеральной совокупности, называется *точечной оценкой*.

Например, $\overline{X}$ – среднее арифметическое – может служить оценкой математического ожидания $M(X)$ генеральной совокупности. В принципе для неизвестного параметра a может существовать много числовых характеристик выборки, которые являются вполне подходящими для того, чтобы служить оценками. Например, среднее арифметическое, медиана, мода могут показаться вполне приемлемыми для оценивания математического ожидания $M(X)$ генеральной совокупности. Чтобы решить, какая из статистик в данном множестве наилучшая, необходимо определить некоторые желаемые свойства таких оценок, т. е. указать условия, которым должны удовлетворять оценки. Такими условиями являются несмещенность, эффективность и состоятельность.

Несмещенность. Если $M(\tilde{a}) = a$, то $\tilde{a}$ называется несмещенной оценкой a.

В других случаях говорят, что оценка смещена.

Несмещенность оценки означает, что при ее использовании в одних случаях может получиться завышение искомого параметра совокупности, в других – занижение. Однако в среднем мы будем «попадать в цель».

Так, например, несмещенной оценкой для математического ожидания $M(X) = a$ случайной величины X является средняя арифметическая $\overline{X} = \tilde{a}$.

Действительно,

$$M(\tilde{a}) = M(\overline{X}) = M\left(\sum_i x_i \cdot \frac{1}{n}\right) = M\left(\frac{1}{n} \cdot \sum_i x_i\right) = \frac{1}{n} \cdot \sum_i M(x_i) =$$

$$= \frac{1}{n} \cdot \sum_i M(x) = M(x) = a,$$

так как результаты выборки x_1, x_2, ..., x_n рассматривают как n независимых случайных величин X_1, X_2, ..., X_n, каждая из которых распределена по тому же закону, что и случайная величина X.

Если существует более одной несмещенной оценки, то выбирают более эффективную, т. е. ту, для которой величина второго момента $M(\tilde{a} - a)^2$ меньше.

Эффективность. Оценка $\tilde{a}_1$ называется более эффективной, чем оценка $\tilde{a}_2$, если $M(\tilde{a}_1 - a)^2 < M(\tilde{a}_2 - a)^2$.

Если обозначить через $b = M(a) - a$ смещение оценки, то $M(\tilde{a} - a)^2 = D(\tilde{a}) + b^2$, так как $M(\tilde{a} - M(\tilde{a})) = 0$, $M(b^2) = b^2$ и, следовательно,

$$M(\tilde{a}-M(\tilde{a})+M(\tilde{a})-a)^2=M((\tilde{a}-M(\tilde{a}))+(M(\tilde{a})-a))^2=$$
$$=M((\tilde{a}-M(\tilde{a}))+b)^2=M(\tilde{a}-M(\tilde{a}))^2+$$
$$+2\cdot b\cdot M(\tilde{a}-M(\tilde{a}))+M(b^2)=D(\tilde{a})+b^2.$$

Поэтому более эффективной оценкой будем считать ту несмещенную оценку, которая имеет меньшую дисперсию.

В частности, средняя арифметическая $\overline{X}=\tilde{a}$ является наиболее эффективной оценкой математического ожидания $M(x)=a$, так как

$$D(\overline{X})=D\left(\sum_i x_i\cdot\frac{1}{n}\right)=D\left(\frac{1}{n}\cdot\sum_i x_i\right)=$$
$$=\frac{1}{n^2}\cdot\sum_i D(x_i)=\frac{1}{n^2}\cdot n\cdot D(x_i)=\frac{1}{n}\cdot D(x_i).$$

Все другие оценки $M(X)$ будут обладать большими дисперсиями. Например, $D\left(\overline{Me}\right)=\frac{\pi}{2}\cdot\frac{1}{n}\cdot D(X)$.

При использовании той или иной оценки желательно, чтобы точность оценивания увеличилась с возрастанием объема производимой выборки. Предельная точность будет достигнута в том случае, когда численное значение оценки совпадает со значением параметра при неограниченном увеличении объема выборки. Такие оценки будем называть *состоятельными.*

Состоятельность. Оценка $\tilde{a}$ называется состоятельной оценкой *a,* если при $n\to\infty$ она сходится по вероятности к *a*, т. е. если $\tilde{a}\xrightarrow[n\to\infty]{p}a$.

Например, средняя арифметическая $\overline{X}=\tilde{a}$ является состоятельной оценкой математического ожидания $M(X)=a$ совокупности, так как согласно закону больших чисел $\overline{X}\xrightarrow[n\to\infty]{p}M(X)$

.

Кроме указанных свойств оценок, существуют и другие, с которыми можно познакомиться в учебнике [13] .

Наконец, при построении оценки $\tilde{a}$ должна использоваться вся информация, содержащаяся в выборке, о неизвестном параметре *a*, т. е. оценка должна быть *достаточной*. Если $\tilde{a}$ – достаточная оценка, то никакая другая оценка не может дать о неизвестном параметре *a* дополнительных сведений.

При выборе оценок следует принимать во внимание перечисленные свойства и учитывать относительную простоту вычислений. Нередко выбирается неэффективная оценка только потому, что ее вычисление намного проще, чем вычисление эффективной оценки. Например, при контроле качества продукции мерой разброса совокупности часто служит выборочный размах, используемый вместо более сложной и более эффективной оценки выборочного стандартного отклонения. Отметим, что при оценивании на основе малого числа наблюдений различие в эффективности оценок невелико.

4.1.1. Статистические ошибки

Выборочные статистики, используемые в качестве точечных оценок каких-либо параметров, как правило, не совпадают с их истинными значениями. Величину отклонения выборочных статистик от истинных значений оцениваемых с их помощью параметров называют статистической ошибкой. Такие отклонения являются случайными величинами, и они изменяются от выборки к выборке. Для измерения статистической ошибки некоторого статистического показателя используют его среднеквадратическое отклонение. Эта величина показывает, насколько сильно варьируют отдельные оценки вокруг параметра генеральной совокупности.

В том случае, если распределение случайной величины X не слишком сильно отличается от нормального распределения, а объем выборки не слишком мал $(n \geq 30)$, то среднеквадратическое отклонение для средней арифметической величины может быть найдено по формуле

$$m_{\bar{x}} = S_{\bar{x}} = \frac{S_x}{\sqrt{n}}.$$

Ниже приведены формулы для вычисления стандартных ошибок некоторых статистических показателей.

Ошибка дисперсии:

$$m_{S^2} = S_{S^2} = \frac{S_x^2}{\sqrt{2 \cdot n}}.$$

Ошибка среднеквадратического отклонения:

$$m_S = S_S = \frac{S_x}{\sqrt{2 \cdot n}}.$$

Ошибка коэффициента вариации:

$$m_V = S_V = \frac{V}{\sqrt{n-1}} \cdot \sqrt{\frac{1}{2} + \left(\frac{V}{100}\right)^2} \approx \frac{V}{\sqrt{2 \cdot n}}.$$

Ошибка коэффициента асимметрии:

$$m_A = S_A = \sqrt{\frac{6 \cdot n \cdot (n-1)}{(n-2) \cdot (n+1) \cdot (n+3)}} \approx \sqrt{\frac{6}{n}}.$$

Ошибка коэффициента эксцесса:

$$m_E = S_E = \sqrt{\frac{24 \cdot n \cdot (n-1)^2}{(n-3) \cdot (n-2) \cdot (n+3) \cdot (n+5)}} \approx \cdot \sqrt{\frac{24}{n}} = 2 \cdot \sqrt{\frac{6}{n}}.$$

Ошибка медианы:

$$m_{Me} = S_{Me} = S_{\bar{x}} \cdot \sqrt{\frac{\pi}{2}}.$$

Чтобы сравнивать ошибки оценки среднего для объектов, имеющих разную размерность, часто используют показатель точности, который представляет собой стандартную ошибку оценки среднего, выраженную в процентах от самой средней величины. Эту статистику можно вычислить по формуле

$$P = \frac{m_{\bar{x}}}{\bar{x}} \cdot 100\%.$$

Рассмотрим пример вычисления стандартных ошибок основных статистических показателей, характеризующих диаметры и высоты деревьев в чистом одновозрастном сосновом древостое (табл. 12 и 13). Необходимые значения статистических показателей возьмем из примеров, приведенных в главе 2.

Ряд диаметров:

$$m_{\bar{x}} = S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{7{,}455}{\sqrt{200}} = 0{,}527;$$

$$m_{S^2} = S_{S^2} = \frac{S_x^2}{\sqrt{2 \cdot n}} = \frac{55{,}58}{\sqrt{200}} = 3{,}93;$$

$$m_S = S_S = \frac{S_x}{\sqrt{2 \cdot n}} = \frac{7{,}455}{\sqrt{2 \cdot 200}} = 0{,}373;$$

$$m_V = S_V = \frac{V}{\sqrt{2 \cdot n}} = \frac{23{,}59}{\sqrt{2 \cdot 200}} = 1{,}18;$$

$$m_A = \sqrt{\frac{6}{n}} = \sqrt{\frac{6}{200}} = 0{,}173;$$

$$m_E = 2 \cdot \sqrt{\frac{6}{n}} = 2 \cdot \sqrt{\frac{6}{200}} = 0{,}346;$$

$$m_{Me} = S_{Me} = S_{\bar{x}} \cdot \sqrt{\frac{\pi}{2}} = 0{,}527 \cdot \sqrt{\frac{3{,}14}{2}} = 0{,}660;$$

$$P = \frac{m_{\bar{x}}}{\bar{x}} \cdot 100\% = \frac{0{,}527}{31{,}60} \cdot 100\% = 1{,}67\%.$$

Ряд высот:

$$m_{\bar{x}} = S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{2{,}299}{\sqrt{200}} = 0{,}163;$$

$$m_{S^2} = S_{S^2} = \frac{S_x^2}{\sqrt{2 \cdot n}} = \frac{5{,}284}{\sqrt{200}} = 0{,}374;$$

$$m_S = S_S = \frac{S_x}{\sqrt{2 \cdot n}} = \frac{2{,}299}{\sqrt{2 \cdot 200}} = 0{,}115;$$

$$m_V = S_V = \frac{V}{\sqrt{2 \cdot n}} = \frac{9{,}270}{\sqrt{2 \cdot 200}} = 0{,}464;$$

$$m_A = \sqrt{\frac{6}{n}} = \sqrt{\frac{6}{200}} = 0{,}173;$$

$$m_E = 2 \cdot \sqrt{\frac{6}{n}} = 2 \cdot \sqrt{\frac{6}{200}} = 0{,}346;$$

$$m_{Me} = S_{Me} = S_{\bar{x}} \cdot \sqrt{\frac{\pi}{2}} = 0{,}163 \cdot \sqrt{\frac{3{,}14}{2}} = 0{,}204;$$

$$P = \frac{m_{\bar{x}}}{\bar{x}} \cdot 100\% = \frac{0{,}163}{24{,}80} \cdot 100\% = 0{,}657\%.$$

4.1.2. Метод максимального правдоподобия

Один из важнейших методов нахождения оценок параметров по данным выборки (метод максимального правдоподобия) был предложен Р. Фишером. Этот метод состоит в том, что для получения оценки неизвестного параметра a нужно найти такое значение $\tilde{a}$, при котором вероятность реализации случайного вектора (X_1, X_2, ..., X_n) (выборки) была бы максимальной. С этой целью «строится» функция, определяющая вероятность получения выборки x_1, x_2, ..., x_n, и находится точки максимума этой функции, которая и является оценкой неизвестного параметра.

Пусть случайный эксперимент описывается непрерывной случайной величиной X, плотность распределения вероятностей $P(x, a)$ которой содержит неизвестный параметр a. Тогда вероятность получения при n независимых наблюдениях величины X выборки x_1, x_2, ..., x_n, всегда равна нулю, так как случайный вектор (X_1, X_2, ..., X_n) с независимыми составляющими $X_i, i = \overline{1,n}$, имеет плотность распределения вероятностей вида

$$L(x_1, x_2, \ldots, x_n; a) = p(x_1; a) \cdot p(x_2; a) \cdot \ldots \cdot p(x_n; a) = \\ = \prod_{i=1}^{n} p(x_i; a). \tag{98}$$

Вероятность отдельного значения непрерывной случайной величины равна нулю. Поэтому рассматривается вероятность попадания выборки в n-мерный параллелепипед с центром (x_1, x_2, ..., x_n) и ребрами $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$:

$$L(x_1, x_2 \ldots x_n; a) \cdot \Delta x_1 \cdot \Delta x_2 \cdot \ldots \cdot \Delta x_n = \\ = p(x_1; a) \cdot p(x_2; a) \cdot \ldots \cdot p(x_n; a) \cdot \Delta x_1 \cdot \Delta x_2 \cdot \ldots \cdot \Delta x_n \tag{99}$$

и находится максимальное значение этой вероятности. Точка максимума вероятности (99) является точкой максимума функции (98).

Функция $L(x_1, x_2 \ldots x_n; a)$ называется *функцией правдоподобия*, а величина a, являющаяся точкой максимума этой функции, – оценкой, полученной методом наибольшего правдоподобия.

Таким образом, функция правдоподобия для непрерывной случайной величины X определяется плотностями вероятностей наблюдаемой выборки $p(x_i; a), i = \overline{1,n}$, где $p(x_i; a)$ – плотность распределения вероятностей определенного вида (плотность вероятностей нормального, показательного распределений и т. д.) при $x = x_i$.

Пусть X – дискретная случайная величина, заданная частотным рядом распределения (табл. 30) с неизвестным параметром – вероятностью появления события $\{X = x_i\}, i = \overline{1,k}$; $p(X = x_i) = p_i(a)$ при фиксированном значении параметра *a*.

Таблица 30. Ряд распределения дискретной случайной величины

x_i	x_1	x_2	…	x_k
m_i	m_1	m_2	…	m_k

Причем $\sum_{i=1}^{k} m_i = n$.

Тогда вероятность того, что составляющие $X_i, i = \overline{1,k}$ случайного вектора (X_1, X_2, ..., X_k) примут значения x_1, x_2, ..., x_k, причем значение x_i встречалось m_i раз, вычисляется по формуле

$$L(x_1, x_2 \ldots x_n; a) = p_1^{m_1}(a) \cdot p_2^{m_2}(a) \cdot \ldots \cdot p_k^{m_k}(a). \qquad (100)$$

Функция, определяемая соотношением (100), называется *функцией правдоподобия* дискретной случайной величины *X*. Величина $\tilde{a}$, являющаяся точкой максимума этой функции, называется оценкой неизвестного параметра *a*, полученной по методу наибольшего правдоподобия. Пусть функция правдоподобия $L(x_1, x_2 \ldots x_n; a)$ дифференцируема по *a* и при любых возможных значениях $x_i, i = \overline{1,n}$ достигает максимума по *a* в интервале возможных значений *a*. Тогда, согласно известным правилам дифференциального исчисления, оценку $\tilde{a}$ неизвестного параметра *a* распределения случайной величины *X* находят, решая уравнение

$$\frac{\partial L(x_1, x_2, \ldots, x_n; a)}{\partial a} = 0 \qquad (101)$$

или систему уравнений

$$\frac{\partial L(x_1, x_2, \ldots, x_n; a_1, a_2, \ldots a_r)}{\partial a_j} = 0, \quad j = \overline{1,r}, \qquad (102)$$

если требуется оценить *r* неизвестных параметров.

Так как точки максимума функций $L(x_1, x_2 \ldots x_n; a)$ и $\ln\left(L(x_1, x_2 \ldots x_n; a)\right)$ совпадают, иногда удобнее вместо уравнений (101) и (102) решать уравнение

$$\frac{\partial \ln\left(L(x_1, x_2, \ldots, x_n; a)\right)}{\partial a} = \frac{1}{L} \frac{\partial L(x_1, x_2, \ldots, x_n; a)}{\partial a} = 0$$

или систему уравнений

$$\frac{\partial \ln\left(L(x_1, x_2, \ldots, x_n; a_1, a_2, \ldots a_r)\right)}{\partial a_j} =$$

$$= \frac{1}{L} \frac{\partial L(x_1, x_2, \ldots, x_n; a_1, a_2, \ldots a_r)}{\partial a_j} = 0,$$

$$j = \overline{1, r},$$

Приведем примеры функций правдоподобия для конкретных распределений случайной величины X.

13. Если X распределена нормально, то для нахождения оценок параметров m и σ используется функция правдоподобия:

$$L = \frac{1}{\left(\sigma \cdot \sqrt{2 \cdot \pi}\right)^n} \exp\left(- \frac{\sum (x_i - m)^2}{2 \cdot \sigma^2} \right),$$

где $x_i, i = \overline{1, n}$ – наблюдаемые значения случайной величины X.

14. Если дискретная случайная величина X имеет биномиальное распределение $p_n(m) = C_n^m \cdot p^m (1-p)^{n-m}$, $i = \overline{0, n}$, то для оценки вероятности p появления события в каждом испытании по результатам наблюдаемых значений (табл. 30) рассматривается функция правдоподобия:

$$L = \left(C_n^{x_1}\right)^{m_1} \cdot \left(C_n^{x_2}\right)^{m_2} \cdot \ldots \cdot \left(C_n^{x_k}\right)^{m_k} \cdot p^{\sum\limits_{i=1}^{k} x_i \cdot m_i} (1-p)^{n^2 - \sum\limits_{i=1}^{k} x_i \cdot m_i}.$$

Использование метода наибольшего правдоподобия приводит иногда к решению сложных систем уравнений, хотя в результате получаем состоятельные (иногда немного смещенные) оценки (смещение оценок устраняется введением поправок). Если существует эффективная оценка, то ее можно найти с помощью этого метода. Оценки наибольшего правдоподобия асимптотически эффективны и имеют асимптотически нормальное распределение.

4.1.3. Метод моментов

Метод моментов заключается в том, что статистические моменты выборки (формулы (64) и (65)) принимаются в качестве оценок для моментов распределения случайной величины X, зависящих от неизвестных параметров. В свою очередь оцениваемые параметры могут быть выражены в виде определенных функций теоретических моментов. Заменяя теоретические моменты их оценками, т. е. статистическими моментами выборочного распределения, получаем оценки параметров. Заметим, что этот метод отличается простотой, хотя иногда приводит к малоэффективным и смещенным оценкам, т. е. оценки, найденные с помощью этого метода, не являются наилучшими. Однако для параметров нормального распределения метод моментов дает эффективные и состоятельные оценки.

4.2. Интервальное оценивание

Мы рассмотрели способы оценки неизвестных параметров закона распределения случайной величины X по данным выборки. Получаемые при этом точечные оценки $\tilde{a}_i$ не совпадают (за исключением редких случаев) с истинным значением неизвестных параметров a_i. Следовательно, всегда имеется некоторая погрешность при замене неизвестного параметра его оценкой, т. е. $|\tilde{a}-a|>0$. Величина погрешности при этом неизвестна, хотя требуется знать, к каким ошибкам может привести замена параметра его точечной оценкой.

Чтобы получить представление о точности и надежности оценки $\tilde{a}$ для параметра a, в математической статистике рассматривают вероятность неравенства $|\tilde{a}-a|<\delta$:

$$P(|\tilde{a}-a|<\delta)=P(-\delta<\tilde{a}-a<\delta)=P(\tilde{a}-\delta<a<\tilde{a}+\delta). \qquad (103)$$

И если эта вероятность близка к единице, т. е. если

$$P(\tilde{a}-\delta<a<\tilde{a}+\delta)=1-\alpha,$$

то диапазон практически возможных значений ошибки, возникающей при замене a на $\tilde{a}$, равен $\pm\delta$. Причем большие по абсолютной величине ошибки появляются с вероятностью $\alpha, \alpha>0$. Чем меньше для данного $\alpha>0$ будет $\delta>0$, тем точнее оценка $\tilde{a}$. Из соотношения (103) видно, что вероятность того, что интервал $]\tilde{a}-\delta;\tilde{a}+\delta[$ со случайными концами накроет неизвестный параметр, равна $p=1-\alpha$. Эта вероятность называется *доверительной вероятностью*.

Доверительный интервал для параметра a – это случайный интервал $]\tilde{a}-\delta;\tilde{a}+\delta[$, определяемый результатами наблюдений, который с заданной доверительной вероятностью $p=1-\alpha$ накрывает неизвестный параметр a.

Граничные точки доверительного интервала называются соответственно *нижним и верхним доверительными пределами.*

Вероятность того, что интервал $]\tilde{a}-\delta; \tilde{a}+\delta[$ не накроет неизвестный параметр $a = 1 - p$, называется *уровнем значимости*.Заданному $p = 1-\alpha$ соответствует не единственный доверительный интервал. Доверительные интервалы могут изменяться от выборки к выборке. Более того, для данной выборки различные методы построения доверительных интервалов могут привести к различным интервалам. Поэтому выработаны определенные правила. Используя их и эффективные оценки неизвестных параметров, получают кратчайшие интервалы для заданной доверительной вероятности $p = 1-\alpha$.

Выбираемый уровень достоверности в определенной степени зависит от необходимой надежности результатов. Практический опыт показывает, что чаще всего используется значение $p = 0{,}95$ (т. е. уровень значимости $\alpha = 0{,}05$), несколько реже $p = 0{,}9$ ($\alpha = 0{,}1$), и $p = 0{,}99$ ($\alpha = 0{,}01$).

Общая процедура получения интервальной оценки следующая.

Записывается вероятностное утверждение вида

$$P\{\delta_1 \le g(a,\tilde{a}) \le \delta_2\} = \int_{\delta_1}^{\delta_2} f(g)dg\,, \tag{104}$$

где *f(g)* – функция плотности вероятности некоторой подходящей статистики *g*. При этом значения δ_1 и δ_2 ищутся обычно с учетом дополнительных условий:

$$P\{g(a,\tilde{a}) < \delta_1\} = P\{g(a,\tilde{a}) > \delta_2\} = \frac{1-p}{2} = \frac{\alpha}{2}\,. \tag{105}$$

Аргумент выражения (104) преобразуется таким образом, чтобы в окончательном виде оцениваемый параметр оказался заключенным между величинами, найденными по выборке. Это и будут границы доверительного интервала (l_1, l_2). Статистика $g(a,\tilde{a})$ выбирается таким образом, чтобы она допускала подобное преобразование и имела известную, лучше табулированную функцию плотности вероятности $f(g)$. Последнее обстоятельство существенно упрощает определение значений δ_1 и δ_2.

В качестве иллюстрации получим интервальную оценку математического ожидания m_x нормальной генеральной совокупности с известной дисперсией σ_x^2. Так как

$$g = \frac{\bar{x} - m_x}{\sigma_x / \sqrt{n}}$$

подчиняется нормированному нормальному распределению (табл. 3 приложения), то с учетом (105) соотношение (104) примет вид

$$P\left\{ -U_{\alpha/2} \le \frac{\bar{x} - m_x}{\sigma_x / \sqrt{n}} \le U_{\alpha/2} \right\} = p\,.$$

Преобразуем аргумент следующим образом:

$$\left[-U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}} - \bar{x} \le -m_x \le U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}} - \bar{x} \right]$$

и далее имеем

$$\left[\bar{x} - U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}} \le m_x \le \bar{x} + U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}} \right]. \qquad (106)$$

Таким образом, мы получили следующие границы доверительного интервала:

$$\hat{l}_1 = \bar{x} - U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}};$$

$$\hat{l}_1 = \bar{x} + U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}},$$

Ширина полученного интервала равна

$$L = \left(\hat{l}_2 - \hat{l}_1\right) = 2 \cdot U_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}}.$$

Рассмотрим пример построения такого доверительного интервала. Предположим, у нас имеется нормально распределенная случайная величина, для которой известно значение дисперсии $\sigma_x^2 = 64$. Произведено 16 независимых наблюдений этой величины, по которым найдено среднее арифметическое значение $\bar{x} = 10{,}4$ в качестве оценки математического ожидания изучаемой случайной величины. Требуется построить для математического ожидания 95%-ный доверительный интервал, т. е. доверительный интервал с p = 0,95.

С помощью (106) получаем

$$\left[10{,}4 - U_{\alpha=0{,}025} \cdot \frac{8}{4} \le m_x \le 10{,}4 + U_{\alpha=0{,}025} \cdot \frac{8}{4} \right].$$

В табл. 3 приложения находим значение $U_{\alpha=0,025}=1,96$. С учетом этой величины вычисляем границы 95%-ного доверительного интервала для математического ожидания:

$$[5,32 \le m_x \le 17,08].$$

Ниже приведены еще несколько доверительных интервалов, построенных аналогичным образом для ряда важных случаев (при условии, что анализируемые случайные величины являются нормальными, а наблюдения независимыми).

Доверительный интервал для параметра нормального распределения m_x при неизвестном параметре σ_x^2 (оценка σ_x^2 вычисляется по материалам той же самой выборки, что и границы доверительного интервала):

$$\bar{x} - t_{\alpha/2,n-1} \cdot \frac{S_x}{\sqrt{n}} \le m_x \le \bar{x} + t_{\alpha/2,n-1} \cdot \frac{S_x}{\sqrt{n}}. \qquad (107)$$

Доверительный интервал для параметра нормального распределения σ_x^2 при известном параметре m_x:

$$\frac{S_x^2 \cdot n}{\chi^2_{\alpha/2,n}} \le \sigma_x^2 \le \frac{S_x^2 \cdot n}{\chi^2_{1-\alpha/2,n}}.$$

Доверительный интервал для параметра нормального распределения σ_x^2 при неизвестном параметре m_x:

$$\frac{S_x^2 \cdot (n-1)}{\chi^2_{\alpha/2,n-1}} \le \sigma_x^2 \le \frac{S_x^2 \cdot (n-1)}{\chi^2_{1-\alpha/2,n-1}}.$$

Доверительный интервал для параметра нормального распределения σ_x при неизвестном параметре m_x:

$$\sqrt{\frac{S_x^2 \cdot (n-1)}{\chi^2_{\alpha/2,n-1}}} \le \sigma_x \le \sqrt{\frac{S_x^2 \cdot (n-1)}{\chi^2_{1-\alpha/2,n-1}}}. \qquad (108)$$

Доверительный интервал для отношения дисперсий двух нормально распределенных случайных величин $\sigma_{x_1}^2 / \sigma_{x_2}^2$ при неизвестных параметрах m_{x_1} и m_{x_2}:

$$\frac{S_{x_1}^2 \cdot (n-1)}{S_{x_2}^2 \cdot F_{n_1-1,n_2-1,\alpha/2}} \le \frac{\sigma_{x_1}^2}{\sigma_{x_2}^2} \le \frac{S_{x_1}^2 \cdot (n-1)}{S_{x_2}^2 \cdot F_{n_2-1,n_1-1,\alpha/2}}.$$

В качестве примера построим доверительные интервалы для параметров нормального распределения. Это распределение имеет два параметра: m_x – математическое ожидание и σ – среднеквадратическое отклонение. В том случае, если точечные оценки параметров m и σ получены на основании выборки, доверительный интервал для математического ожидания строится в соответствии со (107) и требует использования квантилей распределения Стьюдента для уровня значимости $\alpha/2$ и числа степеней свободы $n - 1$ (табл. 5 прил.).

Вычислим доверительные интервалы для параметра нормального распределения m_x по материалам вариационных рядов диаметров и высот. Используя средние арифметические 200 измеренных диаметров и высот, вычисленные ранее ((42) и (43)), а также среднеквадратические отклонения (69) и (72) и исходя из предположения, что диаметры деревьев подчиняются закону нормального распределения, найдем доверительные интервалы, накрывающие параметры m и σ с доверительной вероятностью 0,95. В таблице квантилей распределения Стьюдента 5 в приложении находим $t_{0,025;199} = 1{,}972$. Тогда доверительный интервал для среднего арифметического значения с учетом выражения (107) будет следующим:

диаметры

$$\left]31{,}60 - 1{,}972 \cdot \frac{7{,}455}{\sqrt{200}}; 31{,}60 + 1{,}972 \cdot \frac{7{,}455}{\sqrt{200}}\right[\text{, или }]30{,}560; 32{,}640[;$$

высоты

$$\left]24{,}80 - 1{,}972 \cdot \frac{2{,}299}{\sqrt{200}}; 24{,}80 + 1{,}972 \cdot \frac{2{,}299}{\sqrt{200}}\right[\text{, или }]24{,}479; 25{,}121[.$$

Доверительный интервал для среднеквадратического отклонения σ строится следующим образом. Преобразуем (108) как показано ниже:

$$S_x^2 \cdot \gamma_1 \le \sigma_x \le S_x^2 \cdot \gamma_2, \tag{109}$$

где $\gamma_1 = \sqrt{\dfrac{n-1}{\chi^2_{\alpha/2,n-1}}}$; $\gamma_2 = \sqrt{\dfrac{n-1}{\chi^2_{1-\alpha/2,n-1}}}$; $\chi^2_{\alpha/2,n-1}$ и $\chi^2_{1-\alpha/2,n-1}$ – квантили распределения χ^2, соответствующие числу степеней свободы $\gamma = n - 1$ и уровням значимости $\alpha/2$ и $1 - \alpha/2$. Коэффициенты γ_1 и γ_2 приведены в приложении (табл. 6).

Для того чтобы вычислить границы доверительных интервалов для среднеквадратического отклонения по материалам вариационных рядов диаметров и высот, найдем по табл. 6 приложения коэффициенты $\gamma_1 = 0,912$ и $\gamma_2 = 1,11$ для уровня значимости $\alpha = 0,05$. Используя точечные оценки среднеквадратических отклонений для данных по диаметрам (69) и высотам (72) и с учетом выражения (109), получаем доверительные интервалы для среднеквадратического отклонения:

$$]7,455 \cdot 0,912; 7,455 \cdot 1,11[\text{, или }]6,799; 8,275[\text{ – диаметры;}$$

$$]2,299 \cdot 0,912; 2,299 \cdot 1,11[\text{, или }]2,097; 2,552[\text{ – высоты.}$$

5. СТАТИСТИЧЕСКИЕ ВЫВОДЫ И ПРИНЯТИЕ РЕШЕНИЙ

Выборочные статистики являются случайными величинами. Они будут изменяться случайным образом от одной выборки к другой, варьируя около среднего уровня, соответствующего характеристике генеральной совокупности. Каждая выборочная характеристика, определяемая как функция выборки, имеет соответствующий закон распределения. В связи со сказанным выше возникает задача статистической проверки гипотез относительно числовых характеристик выборки и их законов распределения.

Информация, полученная при обработке выборки из некоторой совокупности, может быть использована для получения выводов (решений) обо всей совокупности. Подобные выводы (решения) называются *статистическими решениями*.

Статистической гипотезой, обозначаемой H, называется любое предположение относительно вида или параметров распределения случайной величины X, которое может быть проверено по результатам выборки.

Предположение, выдвигаемое в отношении вида распределения случайной величины, называется *непараметрической гипотезой.*

Любое предположение, однозначно определяющее распределение выборки, называется простой гипотезой.

Пусть дано $m + 1$ распределение P_0, P_1, ..., P_m и известно, что выборка x_1, x_2, ..., x_n принадлежит одному из этих распределений. Необходимо определить, к какому именно распределению P_i принадлежит выборка. Каждая из гипотез H_i: H_i – выборка принадлежит распределению P_i будет простой. Гипотезы могут формулироваться и о значениях параметров распределения известного вида. Такие гипотезы называются *параметрическими гипотезами.*Если параметрическая гипотеза содержит только одно предположение относительно параметра, то она называется *простой гипотезой*.

Параметрическая гипотеза называется *сложной*, если состоит из конечного или бесконечного числа простых гипотез. При этом одно из таких предположений выбирается в качестве основного и называется *нулевой гипотезой* H_0. Другие предположения или возможности называют *альтернативными гипотезами* H_0, H_1, ….После формулировки статистических гипотез ставится задача их проверки по результатам случайной выборки. Для проверки статистической гипотезы с помощью статистического критерия устанавливается, соответствуют ли взятые из выборки данные выдвинутой гипотезе, т. е. нужно принять или отвергнуть гипотезу.

Статистическим критерием для проверки гипотез H_0, H_1, ..., H_m называют случайную величину $d(X)$, принимающую значения H_i, т. е. это отображение выборочного пространства на множество рассматриваемых гипотез. Если $d(X) = H_k$, то принимают гипотезу H_k.

Так как статистический критерий формулируется по результатам конечной случайной выборки и выборка может быть неудачной, то при проверке статистической гипотезы всегда существует риск принять ложное решение.

Рассмотрим гипотезу H_0 и альтернативную гипотезу H_1 и допустим, что во всех случаях, когда H_0 – истина, действие A предпочтительнее действия B, в противном случае – предпочтительнее действие B. При проверке гипотезы с помощью статистического критерия может возникнуть одна из четырех ситуаций:

1) гипотеза H0 истинна (и поэтому H1 ложна) и предпринимается действие A (принимается истинная гипотеза);

2) гипотеза H_1 истинна (и поэтому H_0 ложна) и предпринимается действие A (принимается ложная гипотеза);

3) гипотеза H_0 истинна (и поэтому H_1 ложна) и предпринимается действие B (принимается ложная гипотеза);

4) гипотеза H_1 истинна (и поэтому H_0 ложна) и предпринимается действие B (принимается истинная гипотеза).

В ситуациях 2 и 3 совершается ошибка. Однако ошибки в этих двух ситуациях качественно отличаются друг от друга.

Существует два типа ошибок.

– *ошибка первого рода*, состоящая в отвержении гипотезы H_0, когда она истинна;

– *ошибка второго рода*, состоящая в принятии гипотезы H_0, когда она ложна.

Число

$$\alpha_i = \alpha_i(\delta) = P_i(\delta(X) \neq H_i),$$

характеризующее вероятность отвержения гипотезы H_i, когда она верна, называют вероятностями ошибок $(i + 1)$-го рода критерия d.

Правильное решение также может быть принято двумя способами (ситуация 1 и 4):

– когда гипотеза H_0 принимается, ибо она верна (не совершается ошибка 1-го рода);

– когда гипотеза H_0 отвергается, ибо она ложна (не совершается ошибка 2-го рода).

Познакомимся с наиболее распространенными критериями.

5.1. Непараметрические критерии

Непараметрические критерии, называемые также критериями согласия, используются для проверки гипотез, выдвигаемых в отношении вида распределения исследуемой случайной величины по материалам выборки. Рассмотрим некоторые из них.

5.1.1. Критерий Пирсона хи-квадрат (χ^2)

Для того чтобы выяснить, подчиняются ли экспериментальные данные какому-либо закону распределения, надо сформулировать статистическую гипотезу в отношении распределения анализируемой случайной величины и затем проверить ее. Рассмотрим процесс проверки непараметрической гипотезы с помощью одного из критериев согласия – критерия Пирсона. На первом этапе следует выдвинуть нулевую гипотезу, состоящую в том, что анализируемый признак подчиняется какому-либо закону распределения. Далее, исходя из предположения о том, что нулевая гипотеза справедлива, следует вычислить статистику χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_i - \tilde{f}_i\right)^2}{\tilde{f}_i}. \qquad (110)$$

Полученное значение сравнивается с квантилем распределения Пирсона χ^2, приведенного в табл. 7 приложения. В качестве параметров распределения используется уровень значимости (обычно используется $\alpha = 0{,}05$) и число степеней свободы:

$$\gamma = k - \rho - 1, \qquad (111)$$

где k – общее число степеней свободы, равное числу слагаемых в сумме из формулы (110); ρ – число параметров теоретической функции распределения, которые оценивались по анализируемым данным.

При вычислении критерия Пирсона следует иметь в виду, что теоретические частоты не должны быть меньше пяти. В том случае, если теоретические частоты оказываются недостаточно большими, следует объединять маленькие классы вместе, формируя таким образом большие классы, имеющие теоретические частоты, превышающие пять.

Рассмотрим процесс проверки непараметрической гипотезы с помощью критерия Пирсона на примере распределений диаметров и высот деревьев в сосновом древостое. Нулевая гипотеза будет заключаться в предположении, что анализируемые случайные величины подчиняются закону нормального распределения. Исходя из такого предположения, вычислим статистику χ^2 для вариационного ряда по диаметрам. Для этого составим вспомогательную табл. 31.

Таблица 31. **Вычисление критерия согласия Пирсона χ^2 (диаметры)**

x_i	f_i		$\tilde{f}_i$		$f_i - \tilde{f}_i$	$(f_i - \tilde{f}_i)^2$	$\frac{(f_i - \tilde{f}_i)^2}{\tilde{f}_i}$
	До укрупнения	После укрупнения	До укрупнения	После укрупнения			
14,45	0	3	3,6	9,6	–6,6	43,56	4,54
17,65	3		6,0				
20,85	11	11	11,6	11,6	–0,6	0,36	0,03
24,05	29	29	21,2	21,2	7,8	60,84	2,87
27,25	39	39	28,8	28,8	10,2	104,04	3,61
30,45	32	32	33,6	33,6	–1,6	2,56	0,08
33,65	33	33	32,8	32,8	0,2	0,04	0,00
36,85	23	23	26,6	26,6	–3,6	12,96	0,49
40,05	10	10	18,0	18,0	–8,0	64,00	3,56
43,25	9	9	10,0	10,0	–1,0	1,00	0,10
46,45	3	11	5,0	7,8	3,2	10,24	1,31
49,65	4		2,0				
52,85	2		0,6				
56,05	2		0,2				
59,25	0		0,0				
Сумма	200	200	200,0	200,0	0,0		16,59

При составлении табл. 31 эмпирические и теоретические частоты вариационного ряда диаметров можно взять из табл. 26. Далее следует объединить интервалы таким образом, чтобы теоретические частоты в укрупненных классах были не меньше пяти. Дальнейшие расчеты (три последние колонки табл. 31) выполняем, используя эмпирические и теоретические частоты, полученные после укрупнения классов. Сумма, образовавшаяся в последней колонке таблицы, и будет статистикой Пирсона χ^2. Теперь, пользуясь табл. 7 приложения, найдем соответствующий квантиль распределения Пирсона χ^2, чтобы при сравнении его с вычисленной статистикой χ^2 проверить нулевую гипотезу. Уровень значимости (вероятность отклонения правильной нулевой гипотезы) возьмем $\alpha = 0{,}05$. С учетом объединения интервалов и того, что мы оценили два параметра нормального распределения (σ и m) по материалам наших экспериментальных данных, вычислим число степеней свободы, пользуясь формулой (111):

$$\gamma = k - \rho - 1 = 10 - 2 - 1 = 7\,.$$

Определив необходимые параметры, найдем квантиль распределения Пирсона $\chi^2_{0,05;7}$ = 14,067 по табл. 7 приложения. Так как вычисленная статистика Пирсона χ^2 = 16,59 превышает табличное значение, то мы отклоняем нулевую гипотезу, т. е. распределение диаметров деревьев в древостое не подчиняется закону нормального распределения.

Аналогичным образом вычислим критерий Пирсона для распределения высот деревьев в древостое (табл. 32).

Таблица 32. **Вычисление критерия согласия Пирсона χ^2 (высоты)**

x_i	f_i		$\tilde{f}_i$		$f_i - \tilde{f}_i$	$\left(f_i - \tilde{f}_i\right)^2$	$\dfrac{\left(f_i - \tilde{f}_i\right)^2}{\tilde{f}_i}$
	До укрупнения	После укрупнения	До укрупнения	После укрупнения			
16,05	0	12	0	6,4	5,6	31,36	4,9
17,05	2		0,2				
18,05	1		0,4				
19,05	4		1,6				
20,05	5		4,2				
21,05	2	2	9,4	9,4	-7,4	54,76	5,83
22,05	13	13	17	17	-4	16	0,94
23,05	25	25	26,2	26,2	-1,2	1,44	0,05
24,05	31	31	32,2	32,2	-1,2	1,44	0,04
25,05	32	32	34,6	34,6	-2,6	6,76	0,2
26,05	43	43	29,4	29,4	13,6	184,96	6,29
27,05	24	24	21,8	21,8	2,2	4,84	0,22
28,05	14	14	12,6	12,6	1,4	1,96	0,16

x_i	f_i		$\tilde{f}_i$		$f_i - \tilde{f}_i$	$\left(f_i - \tilde{f}_i\right)^2$	$\frac{\left(f_i - \tilde{f}_i\right)^2}{\tilde{f}_i}$
	До укрупнения	После укрупнения	До укрупнения	После укрупнения			
29,05	3		6,6				
30,05	1	4	2,6	10,4	-6,4	40,96	3,94
31,05	0		1,2				
Сумма	200	200	200,0	200,0	0,0		22,57

Сравнивая полученную статистику Пирсона χ^2 = 22,57 с квантилем распределения Пирсона $\chi^2_{0,05;7}$ = 14,067, приходим к выводу, что и в данном случае следует отклонить нулевую гипотезу, так как вычисленное значение превышает табличное. Таким образом, и распределение высот деревьев в сосновом древостое не подчиняется закону нормального распределения. В связи с этим вопрос о виде распределения диаметров и высот деревьев в древостое остается открытым. Следовательно, надо выдвигать и проверять с помощью критериев согласия статистические гипотезы в отношении других теоретических распределений до тех пор, пока подходящее распределение не будет найдено. Этот процесс связан с большим количеством вычислительных работ, и поэтому его лучше выполнять с помощью вычислительной техники и специальных статистических программ.

5.1.2. Критерий Колмогорова – Смирнова λ

Для оценки достоверности различий между фактическим и теоретическим распределениями наряду с критерием Пирсона (χ^2) может применяться также и критерий λ, предложенный А. Н. Колмогоровым и Н. В. Смирновым. Этот непараметрический критерий представляет собой максимальную разность ($d_{\max}$) между значениями накопленных частот эмпирического и теоретического рядов распределения (без учета знаков значений d), отнесенную к корню квадратному из объема выборочной совокупности:

$$\lambda = \frac{d_{\max}}{\sqrt{n}},$$

где

$$d_{\max} = \max\left|f_i - \tilde{f}_i\right|.$$

Эту же статистику можно определить с помощью эмпирической (16) и теоретической функций распределения как максимальную разность (D_{max}) между значениями этих функций (без учета знаков значений D), умноженную на квадратный корень из объема выборочной совокупности:

$$\lambda = D_{\max} \cdot \sqrt{n}\,, \tag{112}$$

где

$$D_{\max} = \max\left|F_n(x) - F(x)\right|.$$

Данная статистика имеет при больших n приближенную функцию распределения:

$$K(y) = P\{\lambda < y\} = \sum_{k=-\infty}^{\infty} (-1)^k \cdot e^{-2 \cdot k^2 \cdot y^2}\,.$$

Таким образом, вероятность того, что вычисленная статистика λ окажется больше, чем y, равна $1 - K(y) = \alpha$. Следовательно, для проверки непараметрической гипотезы о виде распределения случайной величины необходимо вычисленную статистику λ сравнить с квантилем распределения Колмогорова λ_α. В том случае, если вычисленная статистика окажется больше, чем табличное значение $\lambda > \lambda_\alpha$, мы отвергаем нулевую гипотезу о виде распределения. В противном случае гипотеза принимается. Задаваясь необходимым уровнем значимости α, найти соответствующий квантиль распределения Колмогорова можно с помощью табл. 4 приложения.

Критерий Колмогорова подразумевает, что параметры распределения, в отношении которого выдвигается нулевая гипотеза, известны. В том случае, если это не так и неизвестные параметры оценивались по данным исследуемой выборки, то критерий Колмогорова становиться приближенным в том смысле, что действительный уровень значимости только приближенно равен заданному уровню. Если при проверке непараметрических гипотез используются оценки параметров распределения, полученные по материалам анализируемой выборки, то критерий Колмогорова показывает лучшее согласие теоретических распределений с экспериментальными данными, чем критерий согласия Пирсона χ^2. В связи с этим в таких случаях при проверке непараметрических гипотез используют несколько больший уровень значимости, чем обычно $\alpha = 0{,}10$–$0{,}20$, или оценивают неизвестные параметры по данным выборок большого объема, параллельных исследуемой.

Рассмотрим пример проверки непараметрической гипотезы с помощью критерия Колмогорова – Смирнова. Для этого воспользуемся данными о распределении диаметров и высот деревьев в сосновом древостое (табл. 12 и 13). Нулевая гипотеза будет заключаться в предположении, что анализируемые случайные величины подчиняются закону нормального распределения. Исходя из такого предположения, вычислим статистику λ для вариационного ряда по диаметрам. Для этого составим вспомогательную табл. 33. Значения эмпирической функции распределения получим делением накопленных частот на объем выборки. Для того чтобы найти теоретические значения функции нормального распределения, следует вычислить значения нормированной случайной величины t, соответствующей верхним границам интервалов:

$$t_i = \frac{x_i + \frac{\eta}{2} - \bar{x}}{S_x},$$

где η – величина интервала.

Таблица 33. **Вычисление критерия согласия Колмогорова λ (диаметры)**

Классы, x_i	Частоты, f_i	Накопленные частоты	Эмпирическая функция распределения, F_n	$t_i^{в}$	$\Phi(t_i^{в})$	$\lvert F_n – F \rvert$
17,65	3	3	0,015	–1,66	0,048	0,033
20,85	11	14	0,070	–1,23	0,106	0,036
24,05	29	43	0,215	–0,80	0,212	0,003
27,25	39	82	0,410	–0,37	0,356	**0,054**
30,45	32	114	0,570	0,06	0,524	0,046
33,65	33	147	0,735	0,49	0,688	0,047
36,85	23	170	0,850	0,92	0,821	0,029
40,05	10	180	0,900	1,35	0,911	0,011
43,25	9	189	0,945	1,78	0,961	0,016
46,45	3	192	0,960	2,21	0,986	0,026
49,65	4	196	0,980	2,64	0,996	0,016
52,85	2	198	0,990	3,07	0,999	0,009
56,05	2	200	1,000	3,49	1,000	0,000
∑	200	–	–	–	–	–

Величина интервала для ряда диаметров, представленного в табл. 12, равна 3,2 (10). Средняя арифметическая величина и среднеквадратическое отклонение для диаметров были определены ранее: $\bar{x} = 31{,}60$ (42) и $S_x = 7{,}455$ (69). Вычисленные значения величины t записаны в пятую колонку табл. 33. Пользуясь этими величинами, в табл. 2 приложения найдем значения функции нормального распределения, соответствующие верхним границам интервалов. Абсолютные величины разностей между эмпирической и теоретической функциями распределения приведены в последней колонке вспомогательной таблицы. Максимальное из этих значений равно 0,054.

Далее с помощью формулы (112) вычислим критерий Колмогорова:

$$\lambda = D_{\max} \cdot \sqrt{n} = 0{,}054 \cdot \sqrt{200} = 0{,}76\,.$$

Теперь в табл. 4 приложений для уровня значимости $\alpha = 0{,}05$ найдем квантиль распределения Колмогорова $\lambda_\alpha = 1{,}36$. Так как вычисленное значение критерия Колмогорова $\lambda = 0{,}76$ меньше, чем табличное $\lambda_\alpha = 1{,}36$, то гипотезу о нормальном распределении диаметров деревьев в древостое мы принимаем.

При проверке той же самой гипотезы с помощью критерия Пирсона в пункте 5.1.1. был сделан совершенно противоположный вывод. Такая ситуация может быть объяснена случайностью (при проверке гипотезы в одном из случаев была сделана ошибка), но скорее всего критерий Колмогорова показал лучшую согласованность с экспериментальными данными из-за того, что оценка параметров нормального распределения была выполнена по материалам той же самой выборки, по которой проверялась нулевая гипотеза.

Попробуем с помощью критерия Колмогорова проверить нулевую гипотезу о нормальном распределении высот в сосновых древостоях по материалам выборки, представленной в табл. 13. Для выполнения вычислений воспользуемся полученными ранее для анализируемых данных значениями: величиной интервала, равной 1,0 (15); средним арифметическим $\bar{x} = 24{,}80$ (43) и среднеквадратическим отклонением $S_x = 2{,}299$ (72). Вспомогательные расчеты выполним в табл. 34.

Таблица 34. **Вычисление критерия согласия Колмогорова λ (высоты)**

Классы, x_i	Частоты, f_i	Накопленные частоты	Эмпирическая функция распределения, F_n	$t_i^{\text{в}}$	$\Phi(t_i^{\text{в}})$	$\lvert F_n - F \rvert$
17,05	2	2	0,010	–3,13	0,001	0,009
18,05	1	3	0,015	–2,69	0,004	0,011
19,05	4	7	0,035	–2,25	0,012	0,023
20,05	5	12	0,060	–1,81	0,035	0,025
21,05	2	14	0,070	–1,37	0,085	0,015
22,05	13	27	0,135	–0,94	0,174	0,039
23,05	25	52	0,260	–0,50	0,309	0,049
24,05	31	83	0,415	–0,06	0,476	0,061
25,05	32	115	0,575	0,38	0,648	0,073
26,05	43	158	0,790	0,82	0,794	0,004
27,05	24	182	0,910	1,26	0,896	0,014
28,05	14	196	0,980	1,70	0,955	0,025
29,05	3	199	0,995	2,14	0,984	0,011
30,05	1	200	1	2,58	0,995	0,005
∑	200	–	–	–	–	–

Максимальная разность между эмпирической и теоретической функциями распределения для высот равна 0,073. Используя это значение, с помощью формулы (112) вычислим критерий Колмогорова:

$$\lambda = D_{\max} \cdot \sqrt{n} = 0{,}073 \cdot \sqrt{200} = 1{,}03\,.$$

Теперь в табл. 4 приложений для уровня значимости $\alpha = 0{,}05$ найдем квантиль распределения Колмогорова $\lambda_\alpha = 1{,}36$. Так как вычисленное значение критерия Колмогорова $\lambda = 1{,}03$ меньше, чем табличное $\lambda_\alpha = 1{,}36$, то и в данном случае гипотезу о нормальном распределении высот деревьев в древостое мы принимаем.

В обоих случаях – и для диаметров, и для высот – критерий Пирсона и критерий Колмогорова дали противоположные результаты. Как уже отмечалось выше, критерий Колмогорова подразумевает, что параметры распределения неизвестны. В рассматриваемых примерах значения параметров были неизвестны и мы оценили их по данным анализируемой выборки. В связи с этим в нашей ситуации критерий Колмогорова является приближенным. Попробуем исправить ситуацию, оценив неизвестные параметры с помощью параллельной выборки. Для этого выберем из совокупности, приведенной в табл. 1 приложения, из которой взяты и анализируемые диаметры и высоты деревьев, еще одну параллельную выборку. По данным этой дополнительной выборки оценим неизвестные параметры нормального распределения для высот и диаметров деревьев в древостое. В качестве параллельной выборки возьмем столбец с номером 14 из табл. 1 приложения. Вспомогательные расчеты, необходимые для оценки неизвестных параметров нормального распределения диаметров деревьев, приведены в табл. 35.

Таблица 35. **Вычисление параметров нормального распределения (диаметры)**

№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	47,0	16,63	276,5569	26	28,5	–1,87	3,4969
2	35,3	4,93	24,3049	27	30,3	–0,07	0,0049
3	36,3	5,93	35,1649	28	33,3	2,93	8,5849
4	22,0	–8,37	70,0569	29	36,5	6,13	37,5769
5	48,5	18,13	328,6969	30	37,6	7,23	52,2729
6	35,3	4,93	24,3049	31	26,5	–3,87	14,9769
7	38,0	7,63	58,2169	32	17,5	–12,87	165,6369
8	22,2	–8,17	66,7489	33	46,4	16,03	256,9609
9	23,3	–7,07	49,9849	34	29	–1,37	1,8769
10	38,5	8,13	66,0969	35	38,5	8,13	66,0969
11	28,3	–2,07	4,2849	36	26,2	–4,17	17,3889
12	24,0	–6,37	40,5769	37	15,6	–14,77	218,1529
13	20,0	–10,37	107,5369	38	37,5	7,13	50,8369
14	25,5	–4,87	23,7169	39	40,1	9,73	94,6729
15	22,0	–8,37	70,0569	40	19	–11,37	129,2769
16	30,5	0,13	0,0169	41	15,5	–14,87	221,1169
17	29,6	–0,77	0,5929	42	26	–4,37	19,0969
18	47,4	17,03	290,0209	43	30,6	0,23	0,0529
19	17,5	–12,87	165,6369	44	29,6	–0,77	0,5929
20	33,0	2,63	6,9169	45	33,3	2,93	8,5849
21	26,0	–4,37	19,0969	46	28	–2,37	5,6169
22	27,0	–3,37	11,3569	47	32,5	2,13	4,5369
23	25,0	–5,37	28,8369	48	40	9,63	92,7369
24	28,0	–2,37	5,6169	49	30,3	–0,07	0,0049
25	37,3	6,93	48,0249	50	22,9	–7,47	55,8009
Итого	767,5	–	1822,4225	–	751,2	–	1525,9545
Всего					1518,7	–	3348,3770

Подставляя суммы из табл. 35 в формулы (18) и (58), вычислим среднюю арифметическую и среднеквадратическое отклонение:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1518,7}{50} = 30,37;$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{3348,3770}{50-1}} = 8,266.$$

Теперь с использование новых оценок параметров нормального распределения, полученных по данным параллельной выборки, вычислим еще раз значение критерия Колмогорова. Вспомогательные вычисления приведены в табл. 36.

Таблица 36. **Вычисление критерия согласия Колмогорова λ (диаметры)**

Классы, x_i	Частоты, f_i	Накопленные частоты	Эмпирическая функция распределения, F_n	$t_i^{\text{в}}$	$\Phi(t_i^{\text{в}})$	$\lvert F_n – F\rvert$
17,65	3	3	0,015	–1,35	0,089	0,074
20,85	11	14	0,070	–0,96	0,169	**0,099**
24,05	29	43	0,215	–0,57	0,284	0,069
27,25	39	82	0,410	–0,18	0,429	0,019
30,45	32	114	0,570	0,20	0,579	0,009
33,65	33	147	0,735	0,59	0,722	0,013
36,85	23	170	0,850	0,98	0,836	0,014
40,05	10	180	0,900	1,36	0,913	0,013
43,25	9	189	0,945	1,75	0,960	0,015
46,45	3	192	0,960	2,14	0,984	0,024
49,65	4	196	0,980	2,53	0,994	0,014
52,85	2	198	0,990	2,91	0,998	0,008
56,05	2	200	1,000	3,30	1	0
∑	200	–	–	–	–	–

При использовании новых оценок параметров максимальная разность между эмпирической и теоретической функциями распределения для диаметров возросла до 0,099. Используя эту величину, определим с помощью формулы (112) значение критерия Колмогорова:

$$\lambda = D_{\max} \cdot \sqrt{n} = 0,099 \cdot \sqrt{200} = 1,40.$$

Сравнивая полученное значение $\lambda = 1{,}40$ с квантилем распределения Колмогорова $\lambda_{0,05} = 1{,}36$, мы отклоняем гипотезу о нормальном распределении, так как вычисленное значение превышает табличное. Таким образом, при использовании параллельной выборки для оценки параметров распределения с помощью критерия Колмогорова был сделан такой же вывод, как и с помощью критерия Пирсона.

Сделаем то же самое и с данными по высотам. Вспомогательные вычисления для определения параметров распределения по данным параллельной выборки (столбец номер 14 из табл. 1 приложения) приведены в табл. 37.

Таблица 37. **Вычисление параметров нормального распределения (высоты)**

№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	27,7	3,55	12,6025	26	25,2	1,05	1,1025
2	22,5	–1,65	2,7225	27	25,2	1,05	1,1025
3	24,6	0,45	0,2025	28	25,2	1,05	1,1025
4	22,4	–1,75	3,0625	29	26,6	2,45	6,0025
5	26,1	1,95	3,8025	30	26,6	2,45	6,0025
6	26,2	2,05	4,2025	31	23,3	–0,85	0,7225
7	25,5	1,35	1,8225	32	18,6	–5,55	30,8025
8	24,0	–0,15	0,0225	33	24,6	0,45	0,2025
9	29,0	4,85	23,5225	34	26,5	2,35	5,5225
10	26,2	2,05	4,2025	35	25,5	1,35	1,8225
11	24,5	0,35	0,1225	36	25,6	1,45	2,1025
12	23,6	–0,55	0,3025	37	19,1	–5,05	25,5025
13	22,6	–1,55	2,4025	38	28,6	4,45	19,8025
14	23,5	–0,65	0,4225	39	27,3	3,15	9,9225
15	22,2	–1,95	3,8025	40	18,0	–6,15	37,8225
16	23,3	–0,85	0,7225	41	19,0	–5,15	26,5225
17	22,5	–1,65	2,7225	42	22,5	–1,65	2,7225
18	27,5	3,35	11,2225	43	23,6	–0,55	0,3025
19	19,6	–4,55	20,7025	44	25,2	1,05	1,1025
20	26,0	1,85	3,4225	45	26,0	1,85	3,4225
21	25,5	1,35	1,8225	46	25,0	0,85	0,7225
22	23,6	–0,55	0,3025	47	19,4	–4,75	22,5625
23	23,6	–0,55	0,3025	48	26,6	2,45	6,0025
24	22,6	–1,55	2,4025	49	25,7	1,55	2,4025
25	26,2	2,05	4,2025	50	17,8	–6,35	40,3225
Итого	611,0	–	111,0425	–	596,7	–	255,6225
Всего					1207,7	–	366,6650

Подставляя необходимые суммы из данной таблицы в формулы (18) и (58), получим

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1207{,}7}{50} = 24{,}15\,;$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{366{,}6650}{50-1}} = 2{,}735\,.$$

Далее, пользуясь новыми средним арифметическим значением и среднеквадратическим отклонением, еще раз выполним вспомогательные вычисления (табл. 38) и определим критерий Колмогорова.

Таблица 38. **Вычисление критерия согласия Колмогорова λ (высоты)**

Классы, x_i	Частоты, f_i	Накопленные частоты	Эмпирическая функция распределения, F_n	$t_i^{\text{в}}$	$\Phi(t_i^{\text{в}})$	$\lvert F_n – F\rvert$
17,05	2	2	0,010	–2,41	0,008	0,002
18,05	1	3	0,015	–2,05	0,020	0,005
19,05	4	7	0,035	–1,68	0,046	0,011
20,05	5	12	0,060	–1,32	0,093	0,033
21,05	2	14	0,070	–0,95	0,171	0,101
22,05	13	27	0,135	–0,59	0,278	0,143
23,05	25	52	0,260	–0,22	0,413	**0,153**
24,05	31	83	0,415	0,15	0,560	0,145
25,05	32	115	0,575	0,51	0,695	0,120
26,05	43	158	0,790	0,88	0,811	0,021
27,05	24	182	0,910	1,24	0,893	0,017
28,05	14	196	0,980	1,61	0,946	0,034
29,05	3	199	0,995	1,97	0,976	0,019
30,05	1	200	1	2,34	0,990	0,010
∑	200					

Подставляя в формулу (112) новую максимальную разность между эмпирической и теоретической функциями распределения, вычислим критерий:

$$\lambda = D_{\max} \cdot \sqrt{n} = 0{,}153 \cdot \sqrt{200} = 2{,}16\,.$$

Вычисленное значение $\lambda = 2{,}16$ превышает табличное $\lambda_{0,05} = 1{,}36$, поэтому мы отклоняем гипотезу о нормальном распределении высот деревьев в чистом сосновом древостое. Таким образом, для высот, так же как и для диаметров, при использовании параллельной выборки для оценки параметров распределения с помощью критериев Колмогорова и Пирсона были сделаны одинаковые выводы в отношении гипотезы о нормальном распределении.

Рассмотренные выше примеры использования критерия Колмогорова показывают, что его целесообразно применять в том случае, когда значения параметров известны или когда они могут быть определены по данным выборок большого объема, параллельных исследуемой.

5.2. Параметрические критерии

Параметрические критерии служат для проверки гипотез, сформулированных о значениях параметров распределения известного вида. Рассмотрим некоторые схемы применения таких критериев.

5.2.1. *F*-критерий Фишера

Перед исследователями довольно часто встает задача сравнения изменчивости двух выборок. Решение такой задачи обычно сводится к проверке нулевой гипотезы о равенстве дисперсий этих выборок $H_0 : \sigma_1^2 = \sigma_2^2$. Для нормально распределенных случайных величин проверить такую гипотезу можно с помощью критерия Фишера. Осуществляется проверка следующим способом. По материалам двух выборок объема n_1 и n_2 вычисляются точечные оценки параметров нормального распределения: $\bar{x}_1$, $\bar{x}_2$, $S_{1,\text{несм}}$ и $S_{2,\text{несм}}$. Далее определяется значение F-критерия Фишера:

$$F = \frac{S^2_{1,\text{несм}}}{S^2_{2,\text{несм}}}, \tag{113}$$

где $S_{1,\text{несм}}$ и $S_{2,\text{несм}}$ – наибольшая и наименьшая дисперсии соответственно. В предположении, что нулевая гипотеза верна, выборочная F-статистика имеет распределение Фишера с $\gamma_1 = n_1 - 1$ и $\gamma_2 = n_2 - 1$ степенями свободы. Таким образом, для проверки нулевой гипотезы надо сравнить вычисленное значение F-критерия (113) с квантилем распределения Фишера для выбранного уровня значимости α с $\gamma_1 = n_1 - 1$ и $\gamma_2 = n_2 - 1$ степенями свободы. Если экспериментальное значение критерия больше табличного $F \geq F_{\alpha/2, n_1-1, n_2-1}$, то нулевая гипотеза о равенстве дисперсий отвергается. Если же вычисленная статистика не превышает соответствующий квантиль F-распределения $F < F_{\alpha/2, n_1-1, n_2-1}$, то у нас нет оснований отвергать гипотезу о равенстве дисперсий анализируемых случайных величин.

Рассмотрим пример сравнения дисперсий диаметров на высоте груди в двух сосновых одновозрастных древостоях. В качестве первой выборки будем использовать диаметры первых 30 деревьев из данных, приведенных в табл. 10. Для второй выборки будем использовать первые 25 диаметров из совокупности, приведенной в табл. 1 приложения. Для выполнения расчетов составим вспомогательную табл. 39.

Сначала по формуле (18) вычислим средние арифметические значения для каждой выборки. Для этого воспользуемся суммами наблюдений из табл. 39:

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1,i}}{n_1} = \frac{749{,}8}{30} = 24{,}99;$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2,i}}{n_2} = \frac{764{,}7}{25} = 30{,}59.$$

Зная средние арифметические значения выборок, можно приступать к вычислению квадратов отклонений наблюдений от средних значений (табл. 39). Найдя сумму этих квадратов отклонений, можно приступать к вычислению выборочных дисперсий с помощью формулы (48):

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2}{n_1 - 1} = \frac{972{,}659}{30 - 1} = 33{,}54;$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2,i} - \bar{x}_2)^2}{n_2 - 1} = \frac{885,8865}{25-1} = 36,91.$$

Теперь мы можем определить величину F-критерия по формуле (113):

$$F = \frac{S_{1,\text{несм}}^2}{S_{2,\text{несм}}^2} = \frac{36,91}{33,54} = 1,10,$$

Таблица 39. **Вычисление критерия Фишера**

Первая выборка				Вторая выборка			
№ дерева	Диаметр	$x_{1,i} - \bar{x}_1$	$(x_{1,i} - \bar{x}_1)^2$	№ дерева	Диаметр	$x_{2,i} - \bar{x}$	$(x_{2,i} - \bar{x}_2)^2$
1	23,2	–1,79	3,2041	1	28,2	–2,39	5,7121
2	30,7	5,71	32,6041	2	30,6	0,01	0,0001
3	28,3	3,31	10,9561	3	20,6	–9,99	99,8001
4	26,9	1,91	3,6481	4	26,7	–3,89	15,1321
5	25,4	0,41	0,1681	5	26,5	–4,09	16,7281
6	30,8	5,81	33,7561	6	27,6	–2,99	8,9401
7	20,1	–4,89	23,9121	7	30,5	–0,09	0,0081
8	29,1	4,11	16,8921	8	26,1	–4,49	20,1601
9	29,3	4,31	18,5761	9	29,9	–0,69	0,4761
10	15,9	–9,09	82,6281	10	21,0	–9,59	91,9681
11	23,1	–1,89	3,5721	11	44,0	13,41	179,8281
12	24,4	–0,59	0,3481	12	38,5	7,91	62,5681
13	25,3	0,31	0,0961	13	31,5	0,91	0,8281
14	24,8	–0,19	0,0361	14	28,6	–1,99	3,9601
15	34,3	9,31	86,6761	15	33,0	2,41	5,8081
16	27,6	2,61	6,8121	16	28,1	–2,49	6,2001
17	24,1	–0,89	0,7921	17	38,6	8,01	64,1601
18	23,7	–1,29	1,6641	18	33,5	2,91	8,4681
19	35,8	10,81	116,8561	19	31,0	0,41	0,1681
20	20,1	–4,89	23,9121	20	23,0	–7,59	57,6081
21	21,6	–3,39	11,4921	21	25,0	–5,59	31,2481
22	17,8	–7,19	51,6961	22	32,0	1,41	1,9881
23	19,1	–5,89	34,6921	23	37,6	7,01	49,1401
24	23,4	–1,59	2,5281	24	43,0	12,41	154,0081
25	28,5	3,51	12,3201	25	29,6	–0,99	0,9801
26	14,4	–10,59	112,1481	–	–	–	–
27	26,9	1,91	3,6481	–	–	–	–
28	14,1	–10,89	118,5921	–	–	–	–

29	23,6	–1,39	1,9321	–	–	–	–
30	37,5	12,51	156,5001	–	–	–	–
$\sum$	749,8		972,6590	$\sum$	764,7		885,8865

Найдем квантиль распределения Фишера с $\gamma_1 = n_1 - 1 = 24$ и $\gamma_2 = n_2 - 1 = 29$ степенями свободы для уровня значимости $\alpha = 0{,}05$, воспользовавшись табл. 8 приложения, $F_{0,05;24;29} = 1{,}90$. Так как вычисленное значение критерия Фишера $F = 1{,}10$ не превышает табличное $F_{0,05;24;29} = 1{,}90$, у нас нет оснований отвергать нулевую гипотезу о равенстве дисперсий двух анализируемых выборок, т. е. мы можем считать, что изменчивость диаметров деревьев в обоих сосновых древостоях одинакова.

5.2.2. *t*-критерий Стьюдента

В начале XIX века английский математик В. Госсет (Стьюдент) обнаружил, что величина

$$t = \frac{\bar{x} - m_0}{S_{\bar{x}}}, \qquad (114)$$

где

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$$ – ошибка средней арифметической величины

подчиняется приведенному ниже закону распределения:

$$f(t) = C \cdot \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}. \qquad (115)$$

Этим можно воспользоваться для проверки нулевой гипотезы H_0: $m = m_0$ о равенстве математического ожидания (m) нормально распределенной случайной величины X какому-нибудь числу m_0. Предполагается, что параметры нормального распределения m и σ неизвестны и по материалам выборки для них получены оценки:

$$\tilde{m} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n};$$

$$\tilde{\sigma} = S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}.$$

Так как случайная величина t имеет распределение Стьюдента с $\gamma = n - 1$ степенями свободы (115), если гипотеза H_0 верна, то проверить нулевую гипотезу H_0: $m = m_0$ против альтернативы H_a: $m \neq m_0$ можно следующим образом. По материалам выборки надо вычислить значение t-критерия (114). Далее, по таблицам распределения Стьюдента (табл. 5 приложения) для выбранного уровня значимости α и числа степеней свободы $\gamma = n - 1$ следует найти квантиль распределения $t_{\alpha/2,n-1}$. Если $|t| \geq t_{\alpha/2,n-1}$, то нулевая гипотеза отклоняется в пользу альтернативной. Если $|t| < t_{\alpha/2,n-1}$, то у нас нет оснований для отклонения нулевой гипотезы.

Довольно часто встает вопрос о равенстве математических ожиданий двух нормально распределенных случайных величин. Рассмотрим наиболее распространенный случай, когда параметры m_1, m_2, σ_1 и σ_2 неизвестны, но предполагается, что $\sigma_1 = \sigma_2$. Тогда для проверки нулевой гипотезы H_0: $m_1 = m_2$ против альтернативной гипотезы H_a: $m_1 \neq m_2$ необходимо вычислить выборочную статистику t следующим образом:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S_{\bar{x}_1 - \bar{x}_2}}. \qquad (116)$$

Ошибку разности средних $S_{\bar{x}_1 - \bar{x}_2}$ определяют по формуле

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \qquad (117)$$

где

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1,i} - \bar{x})^2}{n_1}; \quad S_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2,i} - \bar{x})^2}{n_2}.$$

Если обе выборки имеют одинаковый объем, т. е. $n_1 = n_2 = n$, тогда ошибку разности средних можно вычислить по более простой формуле:

$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{S_1+S_2}{n-1}}\,.$$

Если нулевая гипотеза справедлива, то выборочная статистика t (116) имеет распределение Стьюдента с $\gamma = n_1 + n_2 - 2$ степенями свободы. Таким образом, для проверки нулевой гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин при неизвестных параметрах m_1, m_2, σ_1, σ_2 и одинаковых дисперсиях $\sigma_1 = \sigma_2$ необходимо вычислить по материалам выборки значение t-критерия (116). Затем для выбранного уровня значимости α надо найти в табл. 5 приложения квантиль распределения Стьюдента с $\gamma = n_1 + n_2 - 2$ степенями свободы $t_{\alpha/2, n_1+n_2-1}$. Если вычисленная t-статистика окажется больше табличного значения $|t| \ge t_{\alpha/2, n_1+n_2-1}$, нулевую гипотезу следует отвергнуть. В противном случае гипотеза принимается.

Для примера проверим гипотезу о равенстве математических ожиданий высот деревьев в двух чистых сосновых древостоях. Первая выборка в рассматриваемом примере будет представлена высотами 30 первых деревьев из данных, приведенных в табл. 10. Для второй выборки возьмем высоты первых 25 деревьев из совокупности, представленной в табл. 1 приложения. Прежде чем приступать к проверке гипотезы о равенстве математических ожиданий, следует выяснить одинаковы ли дисперсии в анализируемых выборках. Для проверки гипотезы о равенстве дисперсий высот в двух сосновых древостоях составим вспомогательную табл. 40, в которой выполним все вспомогательные расчеты.

Используя суммы наблюдений из табл. 40, по формуле (18) вычислим средние арифметические значения для каждой выборки:

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1,i}}{n_1} = \frac{651{,}0}{30} = 21{,}70;$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2,i}}{n_2} = \frac{604{,}3}{25} = 24{,}17.$$

Теперь, применяя вычисленные средние арифметические значения выборок, можно вычислить квадраты отклонений наблюдений от средних значений (табл. 40). Используя сумму этих квадратов отклонений, вычислим выборочные дисперсии с помощью формулы (48):

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1,i} - \bar{x}_1)^2}{n_1 - 1} = \frac{145,8000}{30 - 1} = 5,03;$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2,i} - \bar{x}_2)^2}{n_2 - 1} = \frac{106,4305}{25 - 1} = 4,43.$$

Разделив большую из полученных дисперсий на меньшую, получим F-критерий (формула (113)):

$$F = \frac{S^2_{1,\text{несм}}}{S^2_{2,\text{несм}}} = \frac{5,03}{4,43} = 1,14.$$

Таблица 40. **Вычисление критерия Фишера**

Первая выборка				Вторая выборка			
№ дерева	Диаметр	$x_{1,i} - \bar{x}_1$	$(x_{1,i} - \bar{x}_1)^2$	№ дерева	Диаметр	$x_{2,i} - \bar{x}$	$(x_{2,i} - \bar{x}_2)^2$
1	20,5	–1,20	1,4400	1	24,6	0,43	0,1849
2	23,0	1,30	1,6900	2	22,2	–1,97	3,8809
3	23,5	1,80	3,2400	3	21,7	–2,47	6,1009
4	25,5	3,80	14,4400	4	23,7	–0,47	0,2209
5	21,5	–0,20	0,0400	5	23,5	–0,67	0,4489
6	22,5	0,80	0,6400	6	25,6	1,43	2,0449
7	18,5	–3,20	10,2400	7	24,5	0,33	0,1089
8	24,5	2,80	7,8400	8	22,6	–1,57	2,4649
9	23,5	1,80	3,2400	9	22,2	–1,97	3,8809
10	18,5	–3,20	10,2400	10	21,6	–2,57	6,6049
11	21,0	–0,70	0,4900	11	28,6	4,43	19,6249
12	21,0	–0,70	0,4900	12	26,5	2,33	5,4289
13	21,0	–0,70	0,4900	13	25,6	1,43	2,0449
14	22,0	0,30	0,0900	14	21,1	–3,07	9,4249
15	23,0	1,30	1,6900	15	26,6	2,43	5,9049
16	19,5	–2,20	4,8400	16	22,3	–1,87	3,4969
17	20,5	–1,20	1,4400	17	25,1	0,93	0,8649
18	20,5	–1,20	1,4400	18	26,0	1,83	3,3489
19	22,0	0,30	0,0900	19	24,5	0,33	0,1089
20	21,5	–0,20	0,0400	20	23,6	–0,57	0,3249
21	21,5	–0,20	0,0400	21	21,6	–2,57	6,6049

22	21,5	–0,20	0,0400	22	25,0	0,83	0,6889
23	22,5	0,80	0,6400	23	27,0	2,83	8,0089
24	24,5	2,80	7,8400	24	27,0	2,83	8,0089
25	25,5	3,80	14,4400	25	21,6	–2,57	6,6049
26	17,5	–4,20	17,6400	–	–	–	–
27	19,5	–2,20	4,8400	–	–	–	–
28	17,5	–4,20	17,6400	–	–	–	–
29	21,5	–0,20	0,0400	–	–	–	–
30	26,0	4,30	18,4900	–	–	–	–
∑	651,0	–	145,8	∑	604,3	–	106,4305

Квантиль распределения Фишера из табл. 8 приложения с $\gamma_1 = n_1 - 1 = 29$ и $\gamma_2 = n_2$ - 1=24 степенями свободы для уровня значимости α = 0,05 $F_{0,05;24;29} = 2{,}58$ превышает вычисленный критерий Фишера $F = 1{,}10$, следовательно, у нас нет оснований отвергать нулевую гипотезу о равенстве дисперсий двух анализируемых выборок. В связи с этим мы можем воспользоваться t-статистикой (116) для проверки гипотезы о равенстве математических ожиданий высот в двух сосняках. Предварительно необходимо вычислить смещенные оценки дисперсий для обеих выборок по формуле (46):

$$S_1^2 = \frac{\sum_{i=1}^{n_1}\left(x_{1,i} - \bar{x}\right)^2}{n_1} = \frac{145{,}8}{30} = 4{,}86\,;$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2}\left(x_{2,i} - \bar{x}\right)^2}{n_2} = \frac{106{,}4305}{25} = 4{,}26\,.$$

На основании полученных статистик определяем ошибку разности средних $S_{\bar{x}_1 - \bar{x}_2}$ по формуле (117):

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} =$$

$$= \sqrt{\frac{30 \cdot 4{,}86 + 25 \cdot 4{,}26}{30 + 25 - 2} \cdot \left(\frac{1}{30} + \frac{1}{25}\right)} = 0{,}5908\,.$$

Теперь по формуле (116) можно вычислить t-критерий:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}} = \frac{21,70 - 24,17}{0,5908} = -4,181\,.$$

После того как мы вычислили t-статистику, следует для выбранного уровня значимости ($\alpha = 0,05$) найти в табл. 5 приложения квантиль распределения Стьюдента с $\gamma = 30 + 25 - 2 = 53$ степенями свободы $t_{0,025,53} = 2,678$. Так как абсолютная величина вычисленной t-статистики больше табличного значения, то нулевую гипотезу следует отвергнуть. Следовательно, два анализируемых сосновых древостоя имеют разную высоту.

6. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Подавляющее большинство параметров, характеризующих лесные сообщества, в той или иной степени связаны между собой. Нетрудно заметить, например, что у деревьев, имеющих более толстые стволы, как правило, и высоты больше, чем у деревьев с меньшими диаметрами. Для описания зависимостей используются функции, которые ставят в соответствие каждому значению независимой переменной определенное значение зависимой переменной: $y = f(x)$. Однако связь между диаметрами и высотами деревьев в древостое не однозначна. Деревья с одинаковыми диаметрами могут иметь разные высоты, и наоборот, деревья, имеющие одинаковые высоты, по диаметрам могут отличаться. Такая нестрогая зависимость называется *корреляционной*.

Если сопоставить высоты и диаметры деревьев в древостое, то станет ясно, что с увеличением диаметра высоты также увеличиваются. Такую связь называют *положительной*. В некоторых случаях зависимость между случайными величинами имеет противоположный характер. Например, с увеличением возраста древостоев уменьшается их густота (число стволов на гектаре). Такая связь называется *отрицательной*.

Для того чтобы оценить тесноту и характер корреляционной зависимости между случайными величинами, существует ряд показателей связи. Рассмотрим некоторые из них.

Наиболее распространенный показатель связи – это *коэффициент корреляции*. Его можно вычислить по формуле

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{n \cdot S_x \cdot S_y}.$$

1.1 В том случае, если вычисления надо выполнять на основе таблицы распределения (корреляционной решетки), следует воспользоваться следующей формулой:

$$R = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n \cdot S_x \cdot S_y}. \qquad (118)$$

Данный показатель оценивает тесноту связи между случайными величинами в случае линейных зависимостей, однако в природе чаще встречаются нелинейные. В таких случаях коэффициент корреляции не может выразить всю полноту связи. Для нелинейных зависимостей лучше использовать показатель, предложенный Пирсоном, который называется *корреляционным отношением*. Он вычисляется как квадратный корень из отношения межгрупповой дисперсии зависимой случайной величины к ее общей дисперсии. В данном случае группы формируются в пределах интервалов вариационного ряда независимой переменной. Корреляционное отношение можно вычислить с помощью следующей формулы:

$$\eta = \sqrt{\frac{\sum_{i=1}^{k} f_{x,i} \cdot (\bar{y}_{x,i} - \bar{y})^2}{n \cdot S_y^2}}. \tag{119}$$

Стандартные ошибки коэффициента корреляции и корреляционного отношения можно оценить с помощью выражений:

$$S_R = \frac{\sqrt{1-R^2}}{\sqrt{n-2}} \tag{120}$$

и

$$S_\eta = \frac{\sqrt{1-\eta^2}}{\sqrt{n-2}}. \tag{121}$$

По соотношению величины коэффициента корреляции и корреляционного отношения можно сделать вывод о характере связи: прямолинейна она или криволинейна. Чем значительнее корреляционное отношение превышает коэффициент корреляции, тем более криволинейной является эта связь. Для оценки степени криволинейности связи вычисляют меру криволинейности как разницу между квадратами корреляционного отношения и коэффициента корреляции:

$$K = \eta^2 - R^2. \tag{122}$$

Вычислим рассмотренные выше показатели связи для пары случайных величин – диаметры и высоты деревьев в древостое. Для того чтобы выполнить вычисления, составим вспомогательную табл. 41. Подставляя значения сумм из данной таблицы в формулы (118) и (119), получим

$$R = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot S_x \cdot S_y} = \frac{2243{,}22}{200 \cdot 7{,}436 \cdot 2{,}293} = 0{,}6578$$

и

$$\eta^2 = \frac{\sum_{i=1}^{k} f_{x,i} \cdot (\bar{y}_{x,i} - \bar{y})^2}{n \cdot S_y^2} = \frac{619{,}77}{200 \cdot 5{,}258} = 0{,}5894,$$

или

$$\eta = \sqrt{0{,}5894} = 0{,}7677.$$

Теперь, пользуясь выражениями (120) и (121), вычислим стандартные ошибки коэффициента корреляции и корреляционного отношения:

$$S_R = \frac{\sqrt{1 - R^2}}{\sqrt{n-2}} = \frac{\sqrt{1 - 0{,}6578^2}}{\sqrt{200-2}} = \frac{0{,}7532}{14{,}07} = 0{,}05\,353;$$

$$S_\eta = \frac{\sqrt{1 - \eta^2}}{\sqrt{n-2}} = \frac{\sqrt{1 - 0{,}7677^2}}{\sqrt{200-2}} = \frac{0{,}6408}{14{,}07} = 0{,}04\,554.$$

Полученные результаты говорят о том, что между диаметрами и высотами деревьев в древостое существует связь, а тот факт, что корреляционное отношение значительно превышает коэффициент корреляции, показывает нам, что эта зависимость скорее криволинейная, чем прямолинейная. Вычислим, пользуясь формулой (122), меру криволинейности для зависимости высот и диаметров:

$$K = \eta^2 - R^2 = 0{,}7677^2 - 0{,}6578^2 = 0{,}5894 - 0{,}4327 = 0{,}1567.$$

Таблица 41. **Вспомогательная таблица для вычисления коэффициента корреляции и корреляционного отношения**

H \ D	17,65	20,85	24,05	27,25	30,45	33,65	36,85	40,05	43,25	46,45	49,65	52,85	56,05	Сумма	$y_i-\bar{y}$
30,05	–	–	–	1	–	–	–	–	–	–	–	–	–	1	5,25
29,05	–	–	–	1	1	–	–	–	–	–	1	–	–	3	4,25
28,05	–	–	1	1	2	1	3	2		1	2	1	–	14	3,25
27,05	–	–	–	1	3	4	5	3	4	2	–	–	2	24	2,25
26,05	–	–	1	7	4	13	10	3	4	–	1	–	–	43	1,25
25,05	–	–	1	7	7	12	2	1	1	–	–	1	–	32	0,25
24,05	–	–	7	7	12	1	3	1	–	–	–	–	–	31	–0,75
23,05	–	1	8	12	2	2	–	–	–	–	–	–	–	25	–1,75
22,05	–	3	8	1	1	–	–	–	–	–	–	–	–	13	–2,75
21,05	–	1	–	1	–	–	–	–	–	–	–	–	–	2	–3,75
20,05	–	4	1	–	–	–	–	–	–	–	–	–	–	5	–4,75
19,05	1	2	1	–	–	–	–	–	–	–	–	–	–	4	–5,75
18,05	–	–	1	–	–	–	–	–	–	–	–	–	–	1	–6,75
17,05	2	–	–	–	–	–	–	–	–	–	–	–	–	2	–7,75
f_x	3	11	29	39	32	33	23	10	9	3	4	2	2	200	–
$\bar{y}_x$	17,72	20,78	22,95	24,61	25,08	25,63	26,18	26,45	26,38	27,38	27,8	26,55	27,05	–	–
$\bar{y}_x-\bar{y}$	–7,08	–4,02	–1,85	–0,19	0,28	0,83	1,38	1,65	1,58	2,58	3	1,75	2,25	–	–
$f_x\cdot(\bar{y}_x-\bar{y})^2$	150,38	177,76	99,25	1,41	2,51	22,73	43,8	27,23	22,47	19,97	36	6,13	10,13	619,77	–
$x_i-\bar{x}$	–13,95	–10,75	–7,55	–4,35	–1,15	2,05	5,25	8,45	11,65	14,85	18,05	21,25	24,45	–	–
$\sum f_{ij}\cdot(x_i-\bar{x})\cdot(y_j-\bar{y})$	296,44	475,69	405,81	31,54	–10,35	55,86	166,69	139,43	166,01	115,09	216,6	74,38	110,03	2243,22	–

7. РЕГРЕССИОННЫЙ АНАЛИЗ

В предыдущем разделе было установлено, что между диаметрами и высотами деревьев существует связь. Наличие связи между случайными величинами, как правило, ставит перед исследователем задачу построения модели этой связи. При этом один из параметров, характеризующий исследуемый объект, рассматривается как зависимая переменная и строится модель его связи с одним или несколькими параметрами, которые берутся в качестве независимых переменных. В данном случае наличие модели, позволяющей оценивать значения высот деревьев в древостое исходя из их диаметра, может оказать большую практическую пользу, так как трудоемкость измерения высоты растущего дерева значительно выше, чем трудоемкость измерения его диаметра. Легче всего построить графическую модель. Для этого достаточно нанести на график все точки, соответствующие исследуемым объектам, откладывая по оси абсцисс независимую переменную, а по оси ординат – зависимую. Затем, сглаживая нанесенные на график точки, следует провести плавную линию таким образом, чтобы она отражала зависимость между переменными, т. е. проходила в области наибольшего сгущения точек, оставляя примерно одинаковое их количество выше и ниже себя. Однако такой метод не лишен недостатков. Во-первых, он весьма субъективен. Разные люди для одних и тех же данных проведут кривые по-разному. Во-вторых, такие модели можно строить лишь в том случае, если анализируется связь зависимой переменной только с одним независимым параметром. Для моделирования зависимостей между различными параметрами чаще используют математические методы.

7.1. Метод наименьших квадратов

Рассмотрим метод наименьших квадратов, используемый в регрессионном анализе для построения моделей связи. Этот метод позволяет вычислить коэффициенты уравнения регрессии заданного вида таким образом, чтобы сумма квадратов отклонений эмпирических значений зависимой переменной от теоретических значений была наименьшей.

Найдем оценку коэффициентов регрессии методом наименьших квадратов для линейного уравнения вида

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_m \cdot x_m, \qquad (123)$$

где y – зависимая переменная; a_0, a_1, …, a_m – коэффициенты регрессии; x_1, x_2, …, x_m – независимые переменные.

Если вычислить оценки зависимого параметра y для выборки объема n с помощью уравнения регрессии (123), то сумма квадратов отклонений полученных оценок от истинных значений будет равна

$$\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n} (y_i - (a_0 + a_1 \cdot x_{1,i} + a_2 \cdot x_{2,i} + \ldots + a_m \cdot x_{m,i}))^2, \qquad (124)$$

где $\tilde{y}_i$ – оценка параметра y, полученная с помощью уравнения (123) для i-того объекта в выборке объемом n; y_i – значение параметра y для i-того объекта; $x_{1,i}$, $x_{2,i}$, …, $x_{m,i}$ – значения параметров x_1, x_2, …, x_m для i-того объекта.

Метод наименьших квадратов заключается в нахождении значений a_0, a_1, …, a_m, при которых сумма квадратов отклонений вычисленных значений параметра y от истинных, задаваемая выражением (124), будет минимальной. Таким образом, если сумму квадратов отклонений рассматривать как функцию от a_0, a_1, …, a_m, этот метод сводится к задаче нахождения минимума функции нескольких переменных. Для решения этой задачи следует найти частные производные выражения (124) по a_0, a_1, …, a_m и приравнять их нулю. Решение полученной таким образом системы уравнений и даст нам оценку коэффициентов регрессии a_0, a_1, …, a_m методом наименьших квадратов.

Найдем частные производные выражения (124) по a_0, a_1, …, a_m:

$$\frac{\partial\left(\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2\right)}{\partial a_0} =$$

$$= \frac{\partial\left(\sum_{i=1}^{n} (y_i - (a_0 + a_1 \cdot x_{1,i} + \ldots + a_m \cdot x_{m,i}))^2\right)}{\partial a_0} = \qquad (125)$$

$$= \sum_{i=1}^{n} (2 \cdot (y_i - (a_0 + a_1 \cdot x_{1,i} + \ldots + a_m \cdot x_{m,i})) \cdot (-1));$$

$$\frac{\partial\left(\sum_{i=1}^{n}(y_i-\widetilde{y}_i)^2\right)}{\partial a_1}=$$

$$=\frac{\partial\left(\sum_{i=1}^{n}(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))^2\right)}{\partial a_1}= \tag{126}$$

$$=\sum_{i=1}^{n}(2\cdot(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))\cdot(-x_{1,i}));$$

…

$$\frac{\partial\left(\sum_{i=1}^{n}(y_i-\widetilde{y}_i)^2\right)}{\partial a_m}=$$

$$=\frac{\partial\left(\sum_{i=1}^{n}(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))^2\right)}{\partial a_m}= \tag{127}$$

$$=\sum_{i=1}^{n}(2\cdot(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))\cdot(-x_{m,i})).$$

Приравняв полученные производные (125)–(127) к нулю, имеем следующую систему уравнений:

$$\begin{cases}\sum_{i=1}^{n}(2\cdot(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))\cdot(-1))=0;\\ \sum_{i=1}^{n}(2\cdot(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))\cdot(-x_{1,i}))=0;\\ \ldots\\ \sum_{i=1}^{n}(2\cdot(y_i-(a_0+a_1\cdot x_{1,i}+\ldots+a_m\cdot x_{m,i}))\cdot(-x_{m,i}))=0.\end{cases} \tag{128}$$

Разделим обе части каждого из уравнений системы (128) на 2 и представим выражения под знаком суммы в каждом уравнении в виде слагаемых:

$$\begin{cases} \sum_{i=1}^{n}(-y_i + a_0 + a_1 \cdot x_{1,i} + \ldots + a_m \cdot x_{m,i}) = 0; \\ \sum_{i=1}^{n}(-y_i \cdot x_{1,i} + a_0 \cdot x_{1,i} + a_1 \cdot x_{1,i}^2 + \ldots + a_m \cdot x_{m,i} \cdot x_{1,i}) = 0; \\ \ldots \\ \sum_{i=1}^{n}(-y_i \cdot x_{m,i} + a_0 \cdot x_{m,i} + a_1 \cdot x_{1,i} \cdot x_{m,i} + \ldots + a_m \cdot x_{m,i}^2) = 0. \end{cases} \tag{129}$$

Теперь представим левые части уравнений системы (129) в виде отдельных сумм:

$$\begin{cases} -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 \cdot x_{1,i} + \ldots + \sum_{i=1}^{n} a_m \cdot x_{m,i} = 0; \\ -\sum_{i=1}^{n} y_i \cdot x_{1,i} + \sum_{i=1}^{n} a_0 \cdot x_{1,i} + \sum_{i=1}^{n} a_1 \cdot x_{1,i}^2 + \ldots + \sum_{i=1}^{n} a_m \cdot x_{m,i} \cdot x_{1,i} = 0; \\ \ldots \\ -\sum_{i=1}^{n} y_i \cdot x_{m,i} + \sum_{i=1}^{n} a_0 \cdot x_{m,i} + \sum_{i=1}^{n} a_1 \cdot x_{1,i} \cdot x_{m,i} + \ldots + \sum_{i=1}^{n} a_m \cdot x_{m,i}^2 = 0. \end{cases} \tag{130}$$

Перенеся суммы, содержащие y_i, в правые части уравнений системы (130) и учитывая, что

$$\sum_{i=1}^{n} a_0 = n \cdot a_0,$$

получим

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} x_{1,i} + \ldots + a_m \cdot \sum_{i=1}^{n} x_{m,i} = \sum_{i=1}^{n} y_i; \\ a_0 \cdot \sum_{i=1}^{n} x_{1,i} + a_1 \cdot \sum_{i=1}^{n} x_{1,i}^2 + \ldots + a_m \cdot \sum_{i=1}^{n} x_{m,i} \cdot x_{1,i} = \sum_{i=1}^{n} y_i \cdot x_{1,i}; \\ \ldots \\ a_0 \cdot \sum_{i=1}^{n} x_{m,i} + a_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{m,i} + \ldots + a_m \cdot \sum_{i=1}^{n} x_{m,i}^2 = \sum_{i=1}^{n} y_i \cdot x_{m,i}. \end{cases} \tag{131}$$

В результате мы построили систему нормальных уравнений (131), решение которой даст нам оценку коэффициентов регрессии a_0, a_1, …, a_m, полученную методом наименьших квадратов.

7.2. Анализ соответствия регрессионной модели экспериментальным данным

Предсказанные с помощью уравнения регрессии значения зависимой переменной $\tilde{y}_i$ в подавляющем большинстве случаев не совпадают с экспериментальными значениями y_i. Отклонения экспериментальных значений зависимой переменной от теоретических можно представить как разницу между отклонением наблюдаемого значения отклика y_i от общего среднего откликов $\bar{y}$ и отклонением предсказанного значения отклика $\tilde{y}_i$ от того же общего среднего $\bar{y}$:

$$y_i - \tilde{y}_i = y_i - \bar{y} - (\tilde{y}_i - \bar{y}). \tag{132}$$

Выражение (132) можно преобразовать к виду

$$(y_i - \bar{y}) = (\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}_i). \tag{133}$$

Теперь возведем обе части выражения (133) в квадрат и просуммируем от $i = 1$ до n:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}((\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}_i))^2 . \tag{134}$$

Раскрыв квадрат в правой части уравнения (134), получим

$$\begin{gathered}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \\ = \sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(2 \cdot (\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i)) + \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2.\end{gathered} \tag{135}$$

Одно из слагаемых из правой части уравнения (135) равно нулю:

$$2 \cdot \sum_{i=1}^{n}((\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i)) = 0,$$

или

$$\sum_{i=1}^{n}((\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i)) = 0. \tag{136}$$

Докажем это. Так как речь идет об анализе регрессионной модели, коэффициенты которой a_0, a_1, …, a_m получены в результате решения системы нормальных уравнений (131), то все уравнения этой системы мы можем рассматривать как верные равенства и использовать их в доказательстве. Сначала разделим первое уравнение системы (131) на n:

$$a_0 + a_1 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} + \ldots + a_m \cdot \frac{\sum_{i=1}^{n} x_{m,i}}{n} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

Учитывая (18), получим

$$\bar{y} = a_0 + a_1 \cdot \bar{x}_1 + \ldots + a_m \cdot \bar{x}_m. \tag{137}$$

Используя полученное выражение (137) и уравнение (123), преобразуем левую часть выражения (136) следующим образом:

$$\sum_{i=1}^{n} \left((\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i) \right) =$$

$$= \sum_{i=1}^{n} \Big(\left(a_0 + a_1 \cdot x_{1,i} + \ldots + a_m \cdot x_{m,1} - a_0 - a_1 \cdot \bar{x}_1 + \ldots + a_m \cdot \bar{x}_m \right) \times$$

$$\times \left(y_i - a_0 - a_1 \cdot x_{1,i} - \ldots - a_m \cdot x_{m,1} \right) \Big).$$

Теперь приведем подобные члены в первом сомножителе в правой части равенства:

$$\sum_{i=1}^{n} \left((\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i) \right) =$$

$$= \sum_{i=1}^{n} \Big(\left(a_1 \cdot (x_{1,i} - \bar{x}_1) + \ldots + a_m \cdot (x_{m,1} - \bar{x}_m) \right) \times$$

$$\times \left(y_i - a_0 - a_1 \cdot x_{1,i} - \ldots - a_m \cdot x_{m,1} \right) \Big).$$

Раскрывая первые скобки в правой части, получаем следующее равенство:

$$\sum_{i=1}^{n} \left((\tilde{y}_i - \bar{y}) \cdot (y_i - \tilde{y}_i) \right) =$$

$$= \sum_{i=1}^{n} \left(\left(a_1 \cdot (x_{1,i} - \bar{x}_1) \right) \cdot \left(y_i - a_0 - a_1 \cdot x_{1,i} - \ldots - a_m \cdot x_{m,1} \right) \right) +$$

$$\ldots$$

$$+ \sum_{i=1}^{n} \left(\left(a_m \cdot (x_{m,1} - \bar{x}_m) \right) \cdot \left(y_i - a_0 - a_1 \cdot x_{1,i} - \ldots - a_m \cdot x_{m,1} \right) \right).$$

Продолжая раскрывать скобки в правой части, получаем следующее выражение:

$$\sum_{i=1}^{n}\left(\left(\tilde{y}_i-\bar{y}\right)\cdot\left(y_i-\tilde{y}_i\right)\right)=$$
$$=\sum_{i=1}^{n}\left(a_1\cdot x_{1,i}\cdot\left(y_i-a_0-a_1\cdot x_{1,i}-\ldots-a_m\cdot x_{m,1}\right)\right)-$$
$$-\sum_{i=1}^{n}\left(\left(a_1\cdot\bar{x}_1\cdot\left(y_i-a_0-a_1\cdot x_{1,i}-\ldots-a_m\cdot x_{m,1}\right)\right)+\right.$$
$$\ldots$$
$$+\sum_{i=1}^{n}\left(a_m\cdot x_{m,1}\cdot\left(y_i-a_0-a_1\cdot x_{1,i}-\ldots-a_m\cdot x_{m,1}\right)\right)-$$
$$-\sum_{i=1}^{n}\left(a_m\cdot\bar{x}_m\cdot\left(y_i-a_0-a_1\cdot x_{1,i}-\ldots-a_m\cdot x_{m,1}\right)\right).$$

Теперь представим правую часть равенства в виде отдельных сумм:

$$\sum_{i=1}^{n}\left(\left(\tilde{y}_i-\bar{y}\right)\cdot\left(y_i-\tilde{y}_i\right)\right)=$$
$$=a_1\cdot\left(\sum_{i=1}^{n}\left(y_i\cdot x_{1,i}\right)-a_0\cdot\sum_{i=1}^{n}x_{1,i}-a_1\cdot\sum_{i=1}^{n}\left(x_{1,i}^2\right)-\ldots-a_m\cdot\sum_{i=1}^{n}\left(x_{1,i}\cdot x_{m,1}\right)\right)-$$
$$-a_1\cdot\bar{x}_1\cdot\left(\sum_{i=1}^{n}y_i-n\cdot a_0-a_1\cdot\sum_{i=1}^{n}x_{1,i}-\ldots-a_m\cdot\sum_{i=1}^{n}x_{m,1}\right)+$$
$$\ldots$$
$$+a_m\cdot\left(\sum_{i=1}^{n}\left(y_i\cdot x_{m,1}\right)-a_0\cdot\sum_{i=1}^{n}x_{m,1}-a_1\cdot\sum_{i=1}^{n}\left(x_{1,i}\cdot x_{m,1}\right)-\ldots-a_m\cdot\sum_{i=1}^{n}\left(x_{m,1}^2\right)\right)-$$
$$-a_m\cdot\bar{x}_m\cdot\left(\sum_{i=1}^{n}y_i-n\cdot a_0-a_1\cdot\sum_{i=1}^{n}x_{1,i}-\ldots-a_m\cdot\sum_{i=1}^{n}x_{m,1}\right).$$

Используя соответствующие равенства системы (131), преобразуем полученное выражение к виду

$$=a_1\cdot\left(\sum_{i=1}^{n}\left(y_i\cdot x_{1,i}\right)-\sum_{i=1}^{n}\left(y_i\cdot x_{1,i}\right)\right)-a_1\cdot\bar{x}_1\cdot\left(\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}y_i\right)+$$
$$\ldots$$
$$+a_m\cdot\left(\sum_{i=1}^{n}\left(y_i\cdot x_{m,1}\right)-\sum_{i=1}^{n}\left(y_i\cdot x_{m,1}\right)\right)-a_m\cdot\bar{x}_m\cdot\left(\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}y_i\right)=0$$

Каждое слагаемое полученного выражения равно нулю и, следовательно, все выражение также равно нулю, что и требовалось доказать.

Учитывая (136), выражение (135) можно представить в следующем виде:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2. \qquad (138)$$

Так как величина $(y_i - \bar{y})$ является отклонением i-того наблюдения от общего среднего, то левая часть уравнения (138) представляет собой сумму квадратов отклонений относительно среднего наблюдений (SS). Слагаемое $\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2$ из правой части уравнения (138) – это сумма квадратов относительно регрессии, а слагаемое $\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$ – сумма квадратов, обусловленная регрессией. Таким образом, общая сумма квадратов отклонений от среднего состоит из двух частей: суммы квадратов отклонений относительно регрессии и суммы квадратов отклонений, обусловленной регрессией. Первая составляющая общей суммы квадратов отклонений – сумма квадратов отклонений относительно регрессии – обусловлена тем, что не все действительные наблюдения лежат на линии регрессии. Чем больше эта сумма, тем больше разброс экспериментальных данных относительно регрессии, и наоборот, чем меньше эта сумма, тем точнее ложатся точки на линию регрессии. Отсюда следует, что качество регрессионной модели тем выше, чем большая часть общей суммы квадратов отклонений относительно среднего приходится на сумму квадратов отклонений, обусловленную регрессией. Если сумма квадратов, обусловленная регрессией, будет много больше, чем сумма квадратов относительно регрессии, то регрессионная модель будет вполне приемлемой. Кроме того, критерием может служить отношение суммы квадратов, обусловленной регрессией, к сумме квадратов относительно среднего. Если эта величина будет близка к единице, то уравнение регрессии также можно считать удачным.

Каждая сумма квадратов связана с числом степеней свободы. Это число показывает, сколько независимых элементов информации участвует в образовании данной суммы квадратов.

В сумме квадратов отклонений относительно среднего имеется

n–1 независимых элементов. Из n чисел $y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y}$ независимы только n–1, так как $\bar{y}$ определяется по данным выборки y_1, y_2, …, y_n с помощью соотношения $\bar{y} = (y_1 + y_2 +, \ldots, + y_n)/n$ и, следовательно, сумма чисел $y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y}$ равна нулю. Таким образом, только n–1 величина из n чисел $y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y}$ является независимой. Оставшееся значение полностью определяется n–1 независимой величиной и соотношением

$$(y_1 - \bar{y}) + (y_2 - \bar{y}) + \ldots + (y_n - \bar{y}) = 0\,.$$

Теперь определим, сколько независимых элементов информации определяет сумму квадратов, обусловленную регрессией. Эту сумму можно вычислить с помощью формулы

$$\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 = \left(a_1 \cdot \sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right) + \ldots + a_m \cdot \sum_{i=1}^{n}\left(x_{m,1} - \bar{x}_m\right)\right)^2,$$

в которой используется m параметров a_1, a_2, …, a_m, являющихся функциями от y_1, y_2, …, y_n. Следовательно, данная сумма квадратов имеет m степеней свободы.

Сумма квадратов относительно регрессии может быть вычислена следующим образом:

$$\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n}\left(y_i - a_0 - a_1 \cdot x_{1,i} - \ldots - a_m \cdot x_{m,1}\right)^2.$$

При этом из n чисел y_1, y_2, …, y_n имеется n–m–1 независимый элемент информации. Это связано с тем, что m + 1 коэффициент a_0, a_1, …, a_m, используемый при вычислении суммы квадратов относительно регрессии, вычислен по материалам выборки y_1, y_2, …, y_n путем решения системы нормальных уравнений.

Таким образом, число степеней свободы, соответствующее сумме квадратов относительно регрессии, вместе с числом степеней свободы, соответствующим сумме квадратов, обусловленной регрессией, дает число степеней свободы, которое имеется при определении суммы квадратов отклонений относительно среднего:

$$n–1 = (n–m–1) + m. \tag{139}$$

С помощью уравнений (138) и (139) мы можем построить таблицу дисперсионного анализа (табл. 42).

Таблица 42. **Таблица дисперсионного анализа. Основное разложение**

Источник вариации	Число степеней свободы	Суммы квадратов отклонений *SS*	Средние квадраты *MS*
Обусловленный регрессией	m	$\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2$	$MS_R = \frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2}{m}$
Относительно регрессии (остаток)	$n–m–1$	$\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$	$s^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-m-1}$
Общий, скорректированный	$n–1$	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	

Средний квадрат получается при делении каждой суммы квадратов на соответствующее ей число степеней свободы.

Более общая форма таблицы дисперсионного анализа получается при добавлении в таблицу корректирующего фактора для среднего игреков ($SS(b_0)$) (табл. 43).

Средний квадрат относительно регрессии s^2 дает оценку дисперсии относительно регрессии, основанную на $n–m–1$ степенях свободы. Эту дисперсию часто называют остаточной ($\sigma^2_{Y \cdot X}$). Если уравнение регрессии будет оцениваться на основании бесконечно большого количества наблюдений, то квадратный корень из дисперсии относительно регрессии будет представлять собой стандартную ошибку, т. е. ошибку, с которой любое измеренное значение y предсказывается для данного значения x по регрессионному уравнению. Таким образом, стандартную ошибку уравнения регрессии можно вычислить с помощью формулы

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-m-1}}. \tag{140}$$

Таблица 43. **Таблица дисперсионного анализа, включающая *SS*(b_0)**

Источник	Число степеней свободы	*SS*	*MS*

Регрессия$\|b_0$	m	$SS(R\mid b_0 = \sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2$	MS_R
Остаток	n-m-1	$\sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right)^2$	s^2
Общий, скорректированный	n-1	$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$	
Корректирующий фактор (обусловленный b_0)	1	$SS(b_0) = \left(\sum_{i=1}^{n} y_i\right)^2 \Big/ n = n\cdot\bar{y}^2$	
Общий	n	$\sum_{i=1}^{n} y_i^2$	

В связи с тем, что y_i – это случайные величины, любая функция от них также является случайной величиной. Таким образом, MS_R – средний квадрат, обусловленный регрессией, и s^2 – средний квадрат, обусловленный остаточной вариацией, тоже будут случайными величинами. Можно показать, что эти случайные функции имеют следующие средние значения:

$$E(MS_R) = \sigma^2 + \beta_1^2 \cdot \sum (x_{1,i} - \bar{x}_1)^2 + \ldots + \beta_m^2 \cdot \sum (x_{m,i} - \bar{x}_m)^2\ ;$$

$$E(s^2) = \sigma^2,$$

где $E(Z)$ – это среднее или математическое ожидание случайной величины Z.

В предположении, что ошибки ε_i имеют распределение $N(0, \sigma_2)$, можно показать, что если $\beta_1 = \beta_2 = \ldots = \beta_m = 0$, то величины MS_R, и s^2, умноженные на свои числа степеней свободы, следуют распределению Пирсона χ^2 с теми же самыми числами степеней свободы. Так как эти величины независимы, то отношение

$$F = \frac{MS_R}{s^2} \tag{141}$$

подчиняется F-распределению Фишера с m и n–m–1 степенями свободы при условии, что $\beta_1 = \beta_2 = \ldots = \beta_m = 0$. В связи с этим статистику (141) можно использовать как статистический критерий для проверки гипотезы об отсутствии связи между зависимой переменной y и независимыми переменными x_1, x_2, …, x_m. Эта гипотеза принимается, если вычисленное значение критерия Фишера

F меньше, чем квантиль F-распределения Фишера с $\gamma_1 = m$ и $\gamma_2 = n–m–1$ степенями свободы $F_{\alpha, m, n-m-1}$, который можно найти в табл. 8 приложения, и отвергается в противном случае.

Кроме F-критерия Фишера, используемого для проверки гипотезы об отсутствии связи между зависимой и независимыми переменными, для анализа соответствия регрессионной модели экспериментальным данным часто используется R^2-статистика. Эта величина называется *множественным коэффициентом детерминации* и представляет собой квадрат коэффициента корреляции R между y и $\tilde{y}$, который обычно называют *множественным коэффициентом корреляции*. Коэффициент детерминации показывает, какую долю общего разброса относительно среднего значения y объясняет регрессионная модель, и вычисляется как отношение суммы квадратов, обусловленной регрессией, к сумме квадратов относительно среднего:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}. \tag{142}$$

Величина коэффициента детерминации может изменяться от нуля до единицы.

Множественный коэффициент детерминации не учитывает число степеней свободы. В связи с этим данная статистика может использоваться для сравнения между собой уравнений только с одинаковым количеством параметров. Для сравнения уравнений регрессии, построенных на базе наборов параметров разного размера, иногда используют приведенную R^2 статистику. Эта величина учитывает число степеней свободы и определяется следующим образом:

$$R_a^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1}. \tag{143}$$

Рассмотренные выше статистики характеризовали уравнение в целом. Вместе с тем можно проверить значимость оценок каждого коэффициента регрессии в отдельности. Для этого следует проверить нулевую гипотезу, заключающуюся в том, что коэффициент регрессии равен нулю или в более общем случае, любому числу $H_0 : a_j = a_{j,0}$.

Чтобы проверить такую гипотезу, следует вычислить

t-критерий:

$$t = \frac{a_j - a_{j,0}}{s_{a_j}}, \tag{144}$$

где $a_{j,0}$ – число, о равенстве которому коэффициента a_j выдвинута нулевая гипотеза (для проверки значимости оценки коэффициента регрессии принимают $a_{j,0} = 0$), s_{a_j} – оценка стандартного отклонения для регрессионного коэффициента a_j.

Далее абсолютное значение вычисленного t-критерия сравнивают с квантилем t-распределения Стьюдента $t_{n\text{-}m\text{-}1,\ \alpha/2}$ для вероятности $\alpha/2$ и $n–m–1$ степеней свободы, который можно найти в табл. 5 приложения. Если оно превышает табличное, то нулевая гипотеза отвергается, и если она состояла в равенстве коэффициента регрессии нулю, то оценку этого коэффициента следует считать значимой. Если $|t| < t_{n-m-1,\,\alpha/2}$, то нет оснований отвергать нулевую гипотезу и оценку коэффициента регрессии считают значимой.

Для того чтобы вычислить t-критерий, необходимо знать оценку стандартного отклонения s_{a_j} для коэффициента регрессии. Найдем эти значения для случая, когда имеется одна независимая переменная.

Для коэффициента a_1 решение системы нормальных уравнений приводит к выражению

$$a_1 = \frac{\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)\cdot y_i}{\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2} =$$

$$= \frac{(x_{1,1} - \bar{x}_1)\cdot y_1 + (x_{1,2} - \bar{x}_1)\cdot y_2 + \ldots + (x_{1,n} - \bar{x}_1)\cdot y_n}{\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2}.$$

Известно, что дисперсия линейной функции

$$f = b_1\cdot y_1 + b_2\cdot y_2 + \ldots + b_n\cdot y_n$$

может быть выражена через дисперсии составляющих ее слагаемых следующим образом:

$$V(f) = b_1^2\cdot V(y_1) + b_2^2\cdot V(y_2) + \ldots + b_n^2\cdot V(y_n) \tag{145}$$

при условии, что y_1, y_2, …, y_n попарно некоррелированы, а b_1, b_2, …, b_n – константы.

Учитывая сказанное выше, дисперсию для коэффициента a_1 можно вычислить с помощью следующей формулы:

$$V(a_1) =$$
$$= \frac{(x_{1,1} - \bar{x}_1)^2 \cdot V(y_1) + (x_{1,2} - \bar{x}_1)^2 \cdot V(y_2) + \ldots + (x_{1,n} - \bar{x}_1)^2 \cdot V(y_n)}{\left(\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2 \right)^2}.$$

Если $V(y_i) = \sigma^2$, то

$$V(a_1) = \frac{\sum_{i=1}^{n} (x_{1,1} - \bar{x}_1)^2}{\left(\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2 \right)^2} \cdot \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}. \qquad (146)$$

Если σ^2 неизвестна и в предположении, что модель корректна, вместо нее используется оценка s^2 – остаточный средний квадрат, то дисперсию коэффициента регрессии a_1 можно оценить следующим образом:

$$V(a_1) = \frac{s^2}{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}.$$

Учитывая, что стандартное отклонение есть квадратный корень из дисперсии, t-критерий для a_1 можно вычислить по формуле

$$t = (a_1 - a_{1,0}) \cdot \left(\frac{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}{s^2} \right)^{1/2}$$

или

$$t = a_1 \cdot \left(\frac{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}{s^2} \right)^{1/2}, \qquad (147)$$

если проверяется гипотеза о равенстве коэффициента регрессии a_1 нулю.

Теперь найдем стандартное отклонение s_{a_j} для коэффициента

a_0. Сначала запишем выражение (137) для случая одной независимой переменной:

$$\bar{y} = a_0 + a_1 \cdot \bar{x}_1 . \qquad (148)$$

Теперь выразим из уравнения (148) коэффициент a_0:

$$a_0 = \bar{y} - a_1 \cdot \bar{x}_1 . \qquad (149)$$

Преобразуем выражение (149) следующим образом:

$$a_0 = \bar{y} - a_1 \cdot \bar{x}_1 = \frac{1}{n} y_1 + \frac{1}{n} y_2 + \ldots + \frac{1}{n} y_n - \bar{x}_1 \cdot a_1 .$$

Учитывая (145), дисперсию для коэффициента a_0 можно задать выражением

$$V(a_0) = \frac{1}{n^2} V(y_1) + \frac{1}{n^2} V(y_2) + \ldots + \frac{1}{n^2} V(y_n) + \bar{x}_1^2 \cdot V(a_1) . \qquad (150)$$

Используя (146) и учитывая, что $V(y_i) = \sigma^2$, преобразуем выражение (150) следующим образом:

$$V(a_0) = \frac{\sigma^2}{n} + \bar{x}_1^2 \cdot \frac{\sigma^2}{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2} .$$

Теперь вынесем за скобки σ^2 в правой части равенства:

$$V(a_0) = \sigma^2 \cdot \left(\frac{1}{n} + \frac{\bar{x}_1^2}{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2} \right) .$$

Объединяя выражение в скобках в одну дробь, получим

$$V(a_0) = \sigma^2 \cdot \frac{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2 + n \cdot \bar{x}_1^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2} .$$

Теперь раскроем скобки в числителе дроби и представим его в виде отдельных сумм:

$$V(a_0) = \sigma^2 \cdot \frac{\sum_{i=1}^{n} x_{1,i}^2 - 2 \cdot \sum_{i=1}^{n} x_{1,i} \cdot \bar{x}_1 + \sum_{i=1}^{n} \bar{x}_1^2 + n \cdot \bar{x}_1^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2} .$$

После преобразования второго и третьего слагаемого, как показано ниже

$$V(a_0) = \sigma^2 \cdot \frac{\sum_{i=1}^{n} x_{1,i}^2 - 2 \cdot \bar{x}_1 \cdot n \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} + n \cdot \bar{x}_1^2 + n \cdot \bar{x}_1^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2},$$

получим:

$$V(a_0) = \sigma^2 \cdot \frac{\sum_{i=1}^{n} x_{1,i}^2 - 2 \cdot n \cdot \bar{x}_1^2 + 2 \cdot n \cdot \bar{x}_1^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2} = \frac{\sigma^2 \cdot \sum_{i=1}^{n} x_{1,i}^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}.$$

В том случае, когда σ^2 неизвестна, предполагая, что модель корректна, вместо нее можно использовать оценку s^2 – остаточный средний квадрат. Тогда дисперсию коэффициента регрессии a_0 можно оценить следующим образом:

$$V(a_0) = \frac{s^2 \cdot \sum_{i=1}^{n} x_{1,i}^2}{n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}. \tag{151}$$

Теперь, зная дисперсию коэффициента a_0 (151), вычислим t-критерий для проверки гипотезы о равенстве этого коэффициента заданному числу $a_{0,0}$:

$$t = \frac{(a_0 - a_{0,0}) \cdot \left(n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2 \right)^{1/2}}{\left(s^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 \right)^{1/2}}$$

или

$$t = \frac{a_0 \cdot \left(n \cdot \sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2 \right)^{1/2}}{\left(s^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 \right)^{1/2}} \tag{152}$$

для проверки гипотезы о равенстве коэффициента регрессии a_0 нулю.

Мы рассмотрели вариант с одной независимой переменной. С увеличением количества независимых переменных найти стандартное отклонение для коэффициентов регрессии становится значительно труднее. Попробуем это сделать для двух независимых переменных. В рассматриваемом случае система нормальных уравнений (131) примет следующий вид:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} x_{1,i} + a_2 \cdot \sum_{i=1}^{n} x_{2,i} = \sum_{i=1}^{n} y_i; \\ a_0 \cdot \sum_{i=1}^{n} x_{1,i} + a_1 \cdot \sum_{i=1}^{n} x_{1,i}^2 + a_2 \cdot \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} = \sum_{i=1}^{n} y_i \cdot x_{1,i}; \\ a_0 \cdot \sum_{i=1}^{n} x_{2,i} + a_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + a_2 \cdot \sum_{i=1}^{n} x_{2,i}^2 = \sum_{i=1}^{n} y_i \cdot x_{2,i}. \end{cases} \tag{153}$$

Выразим коэффициент a_0 из первого уравнения системы (153)

$$a_0 = \frac{\sum_{i=1}^{n} y_i}{n} - a_1 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} - a_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n}$$

и подставим полученное выражение во второе и третье уравнение системы (153):

$$\begin{cases} \left(\frac{\sum_{i=1}^{n} y_i}{n} - a_1 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} - a_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \right) \cdot \sum_{i=1}^{n} x_{1,i} + \\ + a_1 \cdot \sum_{i=1}^{n} x_{1,i}^2 + a_2 \cdot \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} = \sum_{i=1}^{n} y_i \cdot x_{1,i}; \\ \left(\frac{\sum_{i=1}^{n} y_i}{n} - a_1 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} - a_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \right) \cdot \sum_{i=1}^{n} x_{2,i} + \\ + a_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + a_2 \cdot \sum_{i=1}^{n} x_{2,i}^2 = \sum_{i=1}^{n} y_i \cdot x_{2,i}. \end{cases}$$

После преобразований получим

$$\begin{cases} a_1 \cdot \left(\sum_{i=1}^{n} x_{1,i}^2 - \dfrac{\left(\sum_{i=1}^{n} x_{1,i} \right)^2}{n} \right) + a_2 \cdot \left(\sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i}}{n} \right) = \\ = \sum_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}}{n}; \\ a_1 \cdot \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} x_{1,i} \cdot \sum_{i=1}^{n} x_{2,i}}{n} \right) + a_2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 - \dfrac{\left(\sum_{i=1}^{n} x_{2,i} \right)^2}{n} \right) = \\ = \sum_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{2,i}}{n}. \end{cases} \tag{154}$$

Введем обозначения:

$$\begin{aligned} C_2 &= \sum_{i=1}^{n} x_{1,i}^2 - \frac{\left(\sum_{i=1}^{n} x_{1,i} \right)^2}{n} = n \cdot \left(\overline{x_1^2} - \bar{x}_1^2 \right) = n \cdot S_1^2; \\ C_3 &= \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} - \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i}}{n} = n \cdot S_1 \cdot S_2 \cdot R; \\ C_4 &= \sum_{i=1}^{n} x_{2,i}^2 - \frac{\left(\sum_{i=1}^{n} x_{2,i} \right)^2}{n} = n \cdot \left(\overline{x_2^2} - \bar{x}_2^2 \right) = n \cdot S_2^2. \end{aligned} \tag{155}$$

Тогда система (154) примет вид

$$\begin{cases} a_1 \cdot C_2 + a_2 \cdot C_3 = \sum_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}}{n}; \\ a_1 \cdot C_3 + a_2 \cdot C_4 = \sum_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{2,i}}{n}. \end{cases} \tag{156}$$

Разделим оба уравнения системы (156) на коэффициенты при a_1

$$\begin{cases} a_1 + a_2 \cdot \dfrac{C_3}{C_2} = \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_2}; \\ a_1 + a_2 \cdot \dfrac{C_4}{C_3} = \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_3} \end{cases}$$

и вычтем первое уравнение из второго:

$$a_2 \cdot \left(\frac{C_4}{C_3} - \frac{C_3}{C_2} \right) = \frac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_3} - \frac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_2}. \quad (157)$$

Теперь выразим a_2 из уравнения (157)

$$a_2 = \frac{\dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_3} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_2}}{\dfrac{C_4}{C_3} - \dfrac{C_3}{C_2}}$$

и преобразуем полученное выражение следующим образом:

$$a_2 = \frac{\left(\dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_3} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_2} \right) \cdot C_2 \cdot C_3}{C_2 \cdot C_4 - C_3^2}$$

и далее

$$a_2 = \frac{C_2 \cdot \sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - C_2 \cdot \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n} - C_3 \cdot \sum\limits_{i=1}^{n} y_i \cdot x_{1,i} + C_3 \cdot \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_2 \cdot C_4 - C_3^2}. \quad (158)$$

Теперь перегруппируем слагаемые в числителе выражения (158) следующим образом:

$$a_2 = \frac{\sum_{i=1}^{n} y_i \cdot \left(C_2 \cdot x_{2,i} - C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - C_3 \cdot x_{1,i} + C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} \right)}{C_2 \cdot C_4 - C_3^2}.$$

Учитывая (145) и то, что C_2, C_3 и C_4 – константы, а $V(y_i) = \sigma^2$, дисперсию для коэффициента a_2 можно вычислить с помощью следующей формулы:

$$V(a_2) = \frac{\sum_{i=1}^{n} \sigma^2 \cdot \left(C_2 \cdot x_{2,i} - C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - C_3 \cdot x_{1,i} + C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} \right)^2}{\left(C_2 \cdot C_4 - C_3^2\right)^2}.$$

Вынесем σ^2 за знак суммы и раскроем скобки:

$$V(a_2) = \frac{\sigma^2}{\left(C_2 \cdot C_4 - C_3^2\right)^2} \cdot \sum_{i=1}^{n} \left(\left(C_2 \cdot x_{2,i}\right)^2 - C_2 \cdot x_{2,i} \cdot C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - \right.$$

$$- C_2 \cdot x_{2,i} \cdot C_3 \cdot x_{1,i} + C_2 \cdot x_{2,i} \cdot C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} - C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \cdot C_2 \cdot x_{2,i} +$$

$$+ \left(C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \right)^2 + C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \cdot C_3 \cdot x_{1,i} - C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} \cdot C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} -$$

$$- C_3 \cdot x_{1,i} \cdot C_2 \cdot x_{2,i} + C_3 \cdot x_{1,i} \cdot C_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} + \left(C_3 \cdot x_{1,i}\right)^2 -$$

$$- C_3 \cdot x_{1,i} \cdot C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} + C_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} \cdot C_2 \cdot x_{2,i} -$$

$$-C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}-C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot x_{1,i}+\left(C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\right)^2\Bigg).$$

Теперь приведем подобные члены

$$V(a_2)=\frac{\sigma^2}{\left(C_2\cdot C_4-C_3^2\right)^2}\cdot\sum_{i=1}^{n}\left(\left(C_2\cdot x_{2,i}\right)^2+\left(C_2\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\right)^2+\left(C_3\cdot x_{1,i}\right)^2+\right.$$

$$+\left(C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\right)^2-2\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\cdot C_2\cdot x_{2,i}-2\cdot C_3\cdot x_{1,i}\cdot C_2\cdot x_{2,i}+$$

$$+2\cdot C_3\cdot x_{1,i}\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}+2\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_2\cdot x_{2,i}-$$

$$\left.-2\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}-2\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot x_{1,i}\right)$$

и перегруппируем слагаемые еще раз:

$$V(a_2)=\frac{\sigma^2}{\left(C_2\cdot C_4-C_3^2\right)^2}\cdot\left(C_2^2\cdot\sum_{i=1}^{n}x_{2,i}^2+C_2^2\cdot\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}+C_3^2\cdot\sum_{i=1}^{n}x_{1,i}^2+\right.$$

$$+C_3^2\cdot\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}-2\cdot C_2^2\cdot\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}-2\cdot C_3\cdot C_2\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}+$$

$$+2\cdot C_3\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}+2\cdot C_3\cdot C_2\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}-$$

$$\left. -2\cdot C_3\cdot C_2\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}\cdot\sum\limits_{i=1}^{n}x_{2,i}}{n}-2\cdot C_3^2\cdot\frac{\left(\sum\limits_{i=1}^{n}x_{1,i}\right)^2}{n}\right). \qquad (159)$$

После преобразования выражения (159) получим

$$V(a_2)=\frac{\sigma^2}{\left(C_2\cdot C_4-C_3^2\right)^2}\cdot\left(C_2^2\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\left(\sum\limits_{i=1}^{n}x_{2,i}\right)^2}{n}\right)+\right.$$

$$\left.+C_3^2\cdot\left(\sum_{i=1}^{n}x_{1,i}^2-\frac{\left(\sum\limits_{i=1}^{n}x_{1,i}\right)^2}{n}\right)-2\cdot C_3\cdot C_2\cdot\left(\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}-\frac{\sum\limits_{i=1}^{n}x_{1,i}\cdot\sum\limits_{i=1}^{n}x_{2,i}}{n}\right)\right). \qquad (160)$$

Учитывая (155), выражение (160) преобразуем к виду

$$V(a_2)=\sigma^2\cdot\frac{\left(C_2^2\cdot C_4-C_3^2\cdot C_2\right)}{\left(C_2\cdot C_4-C_3^2\right)^2}=\sigma^2\cdot\frac{C_2\cdot\left(C_2\cdot C_4-C_3^2\right)}{\left(C_2\cdot C_4-C_3^2\right)^2}=$$

$$=\sigma^2\cdot\frac{C_2}{\left(C_2\cdot C_4-C_3^2\right)}.$$

Теперь, используя оценку s^2 – остаточный средний квадрат вместо неизвестной σ^2, в предположении, что модель корректна, получим выражение для вычисления дисперсии коэффициента регрессии a_2:

$$s_{a_2}^2=s^2\cdot\frac{C_2}{C_2\cdot C_4-C_3{}^2}. \qquad (161)$$

Используя (144) и с учетом (161), а также того, что стандартное отклонение есть квадратный корень из дисперсии, t-критерий для a_2 можно вычислить по формуле

$$t=\frac{a_2-a_{2,0}}{s_{a_2}}=\left(a_2-a_{2,0}\right)\cdot\left(\frac{C_2\cdot C_4-C_3^2}{s^2\cdot C_2}\right)^{1/2}$$

или

$$t = \frac{a_2}{s_{a_2}} = a_2 \cdot \left(\frac{C_2 \cdot C_4 - C_3^2}{s^2 \cdot C_2} \right)^{1/2}, \qquad (162)$$

если проверяется гипотеза о равенстве коэффициента регрессии a_2 нулю.

Аналогичным образом получим выражение, позволяющее вычислить t-критерий для коэффициента a_1. Для этого разделим оба уравнения системы (156) на коэффициенты при a_2

$$\begin{cases} a_1 \cdot \dfrac{C_2}{C_3} + a_2 = \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_3}; \\ a_1 \cdot \dfrac{C_3}{C_4} + a_2 = \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_4} \end{cases}$$

и вычтем первое уравнение из второго:

$$a_1 \cdot \left(\frac{C_3}{C_4} - \frac{C_2}{C_3} \right) = \frac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_4} - \frac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_3}. \qquad (163)$$

Теперь выразим a_1 из уравнения (163)

$$a_1 = \frac{\dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{2,i}}{n}}{C_4} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} x_{1,i}}{n}}{C_3}}{\dfrac{C_3}{C_4} - \dfrac{C_2}{C_3}}$$

и преобразуем полученное выражение следующим образом:

$$a_1 = \frac{\left(\frac{\sum_{i=1}^{n} y_i \cdot x_{2,i} - \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{2,i}}{n}}{C_4} - \frac{\sum_{i=1}^{n} y_i \cdot x_{1,i} - \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}}{n}}{C_3} \right) \cdot C_4 \cdot C_3}{C_3^2 - C_2 \cdot C_4}.$$

Далее

$$a_1 = \frac{C_3 \cdot \sum_{i=1}^{n} y_i \cdot x_{2,i} - C_3 \cdot \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{2,i}}{n} - C_4 \cdot \sum_{i=1}^{n} y_i \cdot x_{1,i} + C_4 \cdot \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}}{n}}{C_3^2 - C_2 \cdot C_4}. \quad (164)$$

Теперь перегруппируем слагаемые в числителе выражения (164) следующим образом:

$$a_1 = \frac{\sum_{i=1}^{n} y_i \cdot \left(C_3 \cdot x_{2,i} - C_3 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - C_4 \cdot x_{1,i} + C_4 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} \right)}{C_3^2 - C_2 \cdot C_4}.$$

Учитывая (145) и то, что C_2, C_3 и C_4 константы, а $V(y_i) = \sigma^2$, дисперсию для коэффициента a_1 можно вычислить с помощью следующей формулы:

$$V(a_1) = \frac{\sum_{i=1}^{n} \sigma^2 \cdot \left(C_3 \cdot x_{2,i} - C_3 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - C_4 \cdot x_{1,i} + C_4 \cdot \frac{\sum_{i=1}^{n} x_{1,i}}{n} \right)^2}{\left(C_3^2 - C_2 \cdot C_4\right)^2}.$$

Вынесем σ^2 за знак суммы и раскроем скобки:

$$V(a_1) = \frac{\sigma^2}{\left(C_3^2 - C_2 \cdot C_4\right)^2} \cdot \sum_{i=1}^{n} \left(\left(C_3 \cdot x_{2,i}\right)^2 - C_3 \cdot x_{2,i} \cdot C_3 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{n} - \right.$$

$$-C_3\cdot x_{2,i}\cdot C_4\cdot x_{1,i}+C_3\cdot x_{2,i}\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}-C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\cdot C_3\cdot x_{2,i}+$$

$$+\left(C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\right)^2+C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\cdot C_4\cdot x_{1,i}-C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}-$$

$$-C_4\cdot x_{1,i}\cdot C_3\cdot x_{2,i}+C_4\cdot x_{1,i}\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}+\left(C_4\cdot x_{1,i}\right)^2-$$

$$-C_4\cdot x_{1,i}\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}+C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot x_{2,i}-$$

$$\left.-C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}-C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_4\cdot x_{1,i}+\left(C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\right)^2\right).$$

Теперь приведем подобные члены

$$V(a_1)=\frac{\sigma^2}{\left(C_3^2-C_2\cdot C_4\right)^2}\cdot\sum_{i=1}^{n}\left(\left(C_3\cdot x_{2,i}\right)^2+\left(C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\right)^2+\left(C_4\cdot x_{1,i}\right)^2+\right.$$

$$+\left(C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\right)^2-2\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}\cdot C_3\cdot x_{2,i}-2\cdot C_4\cdot x_{1,i}\cdot C_3\cdot x_{2,i}+$$

$$+2\cdot C_4\cdot x_{1,i}\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}+2\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot x_{2,i}-$$

$$\left.-2\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{2,i}}{n}-2\cdot C_4\cdot\frac{\sum_{i=1}^{n}x_{1,i}}{n}\cdot C_4\cdot x_{1,i}\right)$$

и перегруппируем слагаемые еще раз:

$$V(a_1)=\frac{\sigma^2}{\left(C_3^2-C_2\cdot C_4\right)^2}\cdot\left(C_3^2\cdot\sum_{i=1}^{n}x_{2,i}^2+C_3^2\cdot\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}+C_4^2\cdot\sum_{i=1}^{n}x_{1,i}^2+\right.$$

$$+C_4^2\cdot\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}-2\cdot C_3^2\cdot\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}-2\cdot C_4\cdot C_3\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}+$$

$$+2\cdot C_4\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}+2\cdot C_4\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}-$$

$$\left.-2\cdot C_4\cdot C_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}-2\cdot C_4^2\cdot\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}\right). \qquad (165)$$

После преобразования выражения (165) получим

$$V(a_1)=\frac{\sigma^2}{\left(C_3^2-C_2\cdot C_4\right)^2}\cdot\left(C_3^2\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}\right)+\right.$$

$$\left.+C_4^2\cdot\left(\sum_{i=1}^{n}x_{1,i}^2-\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}\right)-2\cdot C_4\cdot C_3\cdot\left(\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}-\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}\right)\right). \qquad (166)$$

Учитывая (155), выражение (166) преобразуем к виду

$$V(a_1)=\sigma^2\cdot\frac{\left(C_4^2\cdot C_2-C_3^2\cdot C_4\right)}{\left(C_3^2-C_2\cdot C_4\right)^2}=\sigma^2\cdot\frac{C_4\cdot\left(C_2\cdot C_4-C_3^2\right)}{\left(C_3^2-C_2\cdot C_4\right)^2}=$$

$$=\sigma^2\cdot\frac{C_4}{\left(C_2\cdot C_4-C_3^2\right)}.$$

Теперь, используя оценку s^2 – остаточный средний квадрат вместо неизвестной σ^2, в предположении, что модель корректна, получим выражение для вычисления дисперсии коэффициента регрессии a_1:

$$s_{a_1}^2 = s^2 \cdot \frac{C_4}{C_2 \cdot C_4 - C_3^{\ 2}}. \qquad (167)$$

Используя (144) и с учетом (167), а также того, что стандартное отклонение есть квадратный корень из дисперсии, t-критерий для a_2 можно вычислить по формуле

$$t = \frac{a_1 - a_{1,0}}{s_{a_1}} = \left(a_1 - a_{1,0}\right) \cdot \left(\frac{C_2 \cdot C_4 - C_3^2}{s^2 \cdot C_4}\right)^{1/2}$$

или

$$t = \frac{a_1}{s_{a_1}} = a_1 \cdot \left(\frac{C_2 \cdot C_4 - C_3^2}{s^2 \cdot C_4}\right)^{1/2}, \qquad (168)$$

если проверяется гипотеза о равенстве коэффициента регрессии a_1 нулю.

Теперь найдем стандартное отклонение s_{a_j} для коэффициента a_0. Для этого выразим коэффициент a_1 из первого уравнения системы (153)

$$a_1 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_{1,i}} - a_0 \cdot \frac{n}{\sum_{i=1}^{n} x_{1,i}} - a_2 \cdot \frac{\sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}$$

и подставим полученное выражение во второе и третье уравнения данной системы:

$$\begin{cases} a_0 \cdot \sum_{i=1}^{n} x_{1,i} + \left(\dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_{1,i}} - a_0 \cdot \dfrac{n}{\sum_{i=1}^{n} x_{1,i}} - a_2 \cdot \dfrac{\sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) \cdot \sum_{i=1}^{n} x_{1,i}^2 + \\ + a_2 \cdot \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} = \sum_{i=1}^{n} y_i \cdot x_{1,i}; \\ a_0 \cdot \sum_{i=1}^{n} x_{2,i} + \left(\dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_{1,i}} - a_0 \cdot \dfrac{n}{\sum_{i=1}^{n} x_{1,i}} - a_2 \cdot \dfrac{\sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + \\ + a_2 \cdot \sum_{i=1}^{n} x_{2,i}^2 = \sum_{i=1}^{n} y_i \cdot x_{2,i}. \end{cases}$$

Преобразуем полученную систему следующим образом:

$$\begin{cases} a_0 \cdot \left(\sum_{i=1}^{n} x_{1,i} - \dfrac{n \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}} \right) + a_2 \cdot \left(\sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} x_{1,i}^2 \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) = \\ = \sum_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}}; \\ a_0 \cdot \left(\sum_{i=1}^{n} x_{2,i} - \dfrac{n \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) + a_2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 - \dfrac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) = \\ = \sum_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}. \end{cases} \tag{169}$$

Для упрощения выкладок введем обозначения:

$$B_2 = \sum_{i=1}^{n} x_{1,i} - \frac{n \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}}; \tag{170}$$

$$B_3 = \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} - \frac{\sum_{i=1}^{n} x_{1,i}^2 \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}; \qquad (171)$$

$$B_4 = \sum_{i=1}^{n} x_{2,i} - \frac{n \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}; \qquad (172)$$

$$B_5 = \sum_{i=1}^{n} x_{2,i}^2 - \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}. \qquad (173)$$

Используя приведенные выше выражения, запишем систему (169) следующим образом:

$$\begin{cases} a_0 \cdot B_2 + a_2 \cdot B_3 = \sum_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}}; \\ a_0 \cdot B_4 + a_2 \cdot B_5 = \sum_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}. \end{cases}$$

Теперь разделим уравнения полученной системы на коэффициенты при a_2

$$\begin{cases} a_0 \cdot \dfrac{B_2}{B_3} + a_2 = \dfrac{\sum_{i=1}^{n} y_i \cdot x_{1,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}}}{B_3}; \\ a_0 \cdot \dfrac{B_4}{B_5} + a_2 = \dfrac{\sum_{i=1}^{n} y_i \cdot x_{2,i} - \dfrac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}}{B_5} \end{cases}$$

и вычтем первое уравнение из второго:

$$a_0 \cdot \left(\frac{B_4}{B_5} - \frac{B_2}{B_3} \right) = \frac{\sum_{i=1}^{n} y_i \cdot x_{2,i} - \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}}{B_5} -$$

$$- \frac{\sum_{i=1}^{n} y_i \cdot x_{1,i} - \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}}}{B_3}.$$

Выразим из полученного уравнения коэффициент a_0

$$a_0 = B_3 \cdot B_5 \cdot \left(B_3 \cdot \sum_{i=1}^{n} y_i \cdot x_{2,i} - B_3 \cdot \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} - \right.$$

$$\left. - B_5 \cdot \sum_{i=1}^{n} y_i \cdot x_{1,i} + B_5 \cdot \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}} \right) \Bigg/ \left(B_3 \cdot B_4 - B_2 \cdot B_5 \right) \cdot B_3 \cdot B_5$$

и перегруппируем слагаемые следующим образом:

$$a_0 = \sum_{i=1}^{n} y_i \cdot \left(B_3 \cdot x_{2,i} - B_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} - \right.$$

$$\left. - B_5 \cdot x_{1,i} + B_5 \cdot \frac{\sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}} \right) \Bigg/ \left(B_3 \cdot B_4 - B_2 \cdot B_5 \right).$$

Для того чтобы облегчить запись, введем обозначение

$$D = B_5 \cdot \frac{\sum_{i=1}^{n} x_{1,i}^2}{\sum_{i=1}^{n} x_{1,i}} - B_3 \cdot \frac{\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}}. \qquad (174)$$

Тогда выражение для коэффициента a_0 будет выглядеть следующим образом:

$$a_0 = \frac{\sum_{i=1}^{n} y_i \cdot \left(B_3 \cdot x_{2,i} - B_5 \cdot x_{1,i} + D\right)}{B_3 \cdot B_4 - B_2 \cdot B_5}.$$

Используя полученное выражение и учитывая (145) и то, что B_2, B_3, B_4 и B_5 – константы, а $V(y_i) = \sigma^2$, дисперсию для коэффициента a_0 можно вычислить с помощью следующей формулы:

$$V(a_0) = \frac{\sum_{i=1}^{n} \sigma^2 \cdot \left(B_3 \cdot x_{2,i} - B_5 \cdot x_{1,i} + D\right)^2}{\left(B_3 \cdot B_4 - B_2 \cdot B_5\right)^2}. \qquad (175)$$

Вынесем за знак суммы σ^2 и раскроем скобки в числителе выражения (175):

$$V(a_0) = \frac{\sigma^2}{\left(B_3 \cdot B_4 - B_2 \cdot B_5\right)^2} \cdot \sum_{i=1}^{n} \left(B_3^2 \cdot x_{2,i}^2 + B_5^2 \cdot x_{1,i}^2 + D^2 - \right.$$

$$\left. - 2 \cdot B_3 \cdot B_5 \cdot x_{1,i} \cdot x_{2,i} + 2 \cdot B_3 \cdot D \cdot x_{2,i} - 2 \cdot B_5 \cdot x_{1,i} \cdot D\right). \qquad (176)$$

Теперь перегруппируем слагаемые и преобразуем выражение (176), учитывая (174), следующим образом:

$$V(a_0) = \frac{\sigma^2}{\left(B_3 \cdot B_4 - B_2 \cdot B_5\right)^2} \cdot \left(B_3^2 \cdot \sum_{i=1}^{n} x_{2,i}^2 + B_5^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 + n \cdot D^2 - \right.$$

$$\left. - 2 \cdot B_3 \cdot B_5 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + 2 \cdot B_3 \cdot D \cdot \sum_{i=1}^{n} x_{2,i} - 2 \cdot B_5 \cdot D \cdot \sum_{i=1}^{n} x_{1,i} \right) =$$

$$= \frac{\sigma^2}{\left(B_3 \cdot B_4 - B_2 \cdot B_5\right)^2} \cdot \left(B_3^2 \cdot \sum_{i=1}^{n} x_{2,i}^2 + B_5^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 + n \cdot D^2 - \right.$$

$$\left. - 2 \cdot B_5 \cdot \left(B_3 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + D \cdot \sum_{i=1}^{n} x_{1,i} \right) + 2 \cdot B_3 \cdot D \cdot \sum_{i=1}^{n} x_{2,i} \right) =$$

$$= \frac{\sigma^2}{\left(B_3 \cdot B_4 - B_2 \cdot B_5\right)^2} \cdot \left(B_3^2 \cdot \sum_{i=1}^{n} x_{2,i}^2 + B_5^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 + n \cdot D^2 - \right.$$

$$\left. - 2 \cdot B_5^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 + 2 \cdot B_3 \cdot D \cdot \sum_{i=1}^{n} x_{2,i} \right).$$

В полученном выражении приведем подобные члены и подставим вместо D выражение (174):

$$V(a_0)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3^2\cdot\sum_{i=1}^{n}x_{2,i}^2+n\cdot D^2-B_5^2\cdot\sum_{i=1}^{n}x_{1,i}^2+2\cdot B_3\cdot D\cdot\sum_{i=1}^{n}x_{2,i}\right)=$$

$$=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3^2\cdot\sum_{i=1}^{n}x_{2,i}^2+n\cdot\left(B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}-B_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right)^2-\right.$$

$$\left.-B_5^2\cdot\sum_{i=1}^{n}x_{1,i}^2+2\cdot B_3\cdot\left(B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}-B_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right)\cdot\sum_{i=1}^{n}x_{2,i}\right).\qquad(177)$$

Далее раскроем скобки в правой части равенства (177):

$$V(a_0)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3^2\cdot\sum_{i=1}^{n}x_{2,i}^2+n\cdot B_5^2\cdot\frac{\left(\sum_{i=1}^{n}x_{1,i}^2\right)^2}{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}+n\cdot B_3^2\cdot\frac{\left(\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}\right)^2}{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}-\right.$$

$$-2\cdot n\cdot B_3\cdot B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}-$$

$$\left.-B_5^2\cdot\sum_{i=1}^{n}x_{1,i}^2+2\cdot B_3\cdot B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2\cdot\sum_{i=1}^{n}x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}-2\cdot B_3^2\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}\cdot\sum_{i=1}^{n}x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right).\qquad(178)$$

Учитывая (170)–(173), преобразуем выражение (178) следующим образом:

$$V(a_0)=\frac{\sigma^2}{(B_3\cdot B_4-B_2\cdot B_5)^2}\cdot\left(B_3^2\cdot\left(\sum_{i=1}^n x_{2,i}^2-\frac{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}\cdot\sum_{i=1}^n x_{2,i}}{\sum_{i=1}^n x_{1,i}}\right)-\right.$$

$$-B_5^2\cdot\frac{\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}}\cdot\left(\sum_{i=1}^n x_{1,i}-\frac{n\cdot\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}}\right)-$$

$$-B_3^2\cdot\frac{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^n x_{1,i}}\cdot\left(\sum_{i=1}^n x_{2,i}-\frac{n\cdot\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^n x_{1,i}}\right)+$$

$$\left.+2\cdot B_3\cdot B_5\cdot\frac{\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}}\cdot\left(\sum_{i=1}^n x_{2,i}-\frac{n\cdot\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^n x_{1,i}}\right)\right)=$$

$$=\frac{\sigma^2}{(B_3\cdot B_4-B_2\cdot B_5)^2}\cdot\left(B_3^2\cdot B_5-B_2\cdot B_5^2\cdot\frac{\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}}-\right.$$

$$\left.-B_3^2\cdot B_4\cdot\frac{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^n x_{1,i}}+2\cdot B_3\cdot B_4\cdot B_5\cdot\frac{\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}}\right).$$

Далее с учетом (171) и (172) получим

$$V(a_0)=\frac{\sigma^2}{(B_3\cdot B_4-B_2\cdot B_5)^2}\times$$

$$\times\left(B_3\cdot\frac{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^n x_{1,i}}\cdot\left(B_3\cdot B_5\,\frac{\sum_{i=1}^n x_{1,i}}{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}-B_3\cdot B_4+B_4\cdot B_5\cdot\frac{\sum_{i=1}^n x_{1,i}^2}{\sum_{i=1}^n x_{1,i}\cdot x_{2,i}}\right)+\right.$$

$$+B_5\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}^2}{\sum\limits_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)\Bigg)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}}\cdot\left(B_5\cdot\left(\frac{\sum\limits_{i=1}^{n}x_{1,i}}{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}\cdot\left(\sum_{i=1}^{n}x_{2,i}\cdot x_{1,i}-\frac{\sum\limits_{i=1}^{n}x_{1,i}^2\cdot\sum\limits_{i=1}^{n}x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}}\right)+\right.\right.\right.$$

$$\left.\left.\left.+\frac{\sum\limits_{i=1}^{n}x_{1,i}^2}{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}\cdot\left(\sum_{i=1}^{n}x_{2,i}-\frac{n\cdot\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}}\right)\right)-B_3\cdot B_4\right)+\right.$$

$$\left.+B_5\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}^2}{\sum\limits_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)\right).$$

Теперь раскроем скобки в правой части полученного равенства и приведем подобные члены:

$$V\left(a_0\right)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}}\cdot\left(B_5\cdot\left(\sum_{i=1}^{n}x_{1,i}-\frac{\sum\limits_{i=1}^{n}x_{1,i}^2\cdot\sum\limits_{i=1}^{n}x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}+\right.\right.\right.$$

$$\left.\left.\left.+\frac{\sum\limits_{i=1}^{n}x_{1,i}^2\cdot\sum\limits_{i=1}^{n}x_{2,i}}{\sum\limits_{i=1}^{n}x_{1,i}\cdot x_{2,i}}-\frac{n\cdot\sum\limits_{i=1}^{n}x_{1,i}^2}{\sum\limits_{i=1}^{n}x_{1,i}}\right)-B_3\cdot B_4\right)+\right.$$

$$\left.+B_5\cdot\frac{\sum\limits_{i=1}^{n}x_{1,i}^2}{\sum\limits_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)\right)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\cdot\left(B_5\cdot\left(\sum_{i=1}^{n}x_{1,i}-\frac{n\cdot\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}\right)-B_3\cdot B_4\right)+\right.$$

$$\left.+B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)\right).$$

Далее с учетом (170) приходим к выражению

$$V(a_0)=\frac{\sigma^2}{\left(B_3\cdot B_4-B_2\cdot B_5\right)^2}\times$$

$$\times\left(B_5\cdot\frac{\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)-B_3\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)\right)=$$

$$=\frac{\sigma^2\cdot\left(B_5\cdot\sum_{i=1}^{n}x_{1,i}^2-B_3\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}\right)}{\sum_{i=1}^{n}x_{1,i}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)} \qquad (179)$$

С учетом (170)–(173) преобразуем знаменатель полученного выражения следующим образом:

$$\sum_{i=1}^{n}x_{1,i}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)=\sum_{i=1}^{n}x_{1,i}\times$$

$$\times\left(\left(\sum_{i=1}^{n}x_{2,i}\cdot x_{1,i}-\frac{\sum_{i=1}^{n}x_{1,i}^2\cdot\sum_{i=1}^{n}x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right)\cdot\left(\sum_{i=1}^{n}x_{2,i}-\frac{n\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right)-\right.$$

$$\left.-\left(\sum_{i=1}^{n}x_{1,i}-\frac{n\cdot\sum_{i=1}^{n}x_{1,i}^2}{\sum_{i=1}^{n}x_{1,i}}\right)\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\sum_{i=1}^{n}x_{2,i}\cdot\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}}{\sum_{i=1}^{n}x_{1,i}}\right)\right).$$

Теперь раскроем скобки в правой части равенства:

$$\sum_{i=1}^{n} x_{1,i} \cdot \left(B_3 \cdot B_4 - B_2 \cdot B_5\right) = \sum_{i=1}^{n} x_{1,i} \times$$

$$\times \left(\sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} \cdot \sum_{i=1}^{n} x_{2,i} - \sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} \cdot \frac{n \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} - \right.$$

$$- \frac{\sum_{i=1}^{n} x_{1,i}^{2} \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \cdot \sum_{i=1}^{n} x_{2,i} + \frac{\sum_{i=1}^{n} x_{1,i}^{2} \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \cdot \frac{n \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} -$$

$$- \sum_{i=1}^{n} x_{1,i} \cdot \sum_{i=1}^{n} x_{2,i}^{2} + \sum_{i=1}^{n} x_{1,i} \cdot \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} +$$

$$\left. + \frac{n \cdot \sum_{i=1}^{n} x_{1,i}^{2}}{\sum_{i=1}^{n} x_{1,i}} \cdot \sum_{i=1}^{n} x_{2,i}^{2} - \frac{n \cdot \sum_{i=1}^{n} x_{1,i}^{2}}{\sum_{i=1}^{n} x_{1,i}} \cdot \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right).$$

Далее умножим слагаемые в скобках в правой части равенства на $\sum_{i=1}^{n} x_{1,i}$, вынесем n за скобки и перегруппируем слагаемые, как показано ниже:

$$\sum_{i=1}^{n} x_{1,i} \cdot \left(B_3 \cdot B_4 - B_2 \cdot B_5\right) =$$

$$= n \cdot \left(\sum_{i=1}^{n} x_{1,i}^{2} \cdot \sum_{i=1}^{n} x_{2,i}^{2} - \frac{\sum_{i=1}^{n} x_{1,i}^{2} \cdot \left(\sum_{i=1}^{n} x_{2,i} \right)^{2}}{n} - \right.$$

$$- \frac{\left(\sum_{i=1}^{n} x_{1,i} \right)^{2} \cdot \sum_{i=1}^{n} x_{2,i}^{2}}{n} + \frac{\left(\sum_{i=1}^{n} x_{1,i} \right)^{2} \cdot \left(\sum_{i=1}^{n} x_{2,i} \right)^{2}}{n^{2}} - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^{2} +$$

$$+2\cdot\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}\cdot\sum_{i=1}^{n}x_{2,i}\cdot x_{1,i}}{n}-\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2\cdot\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n^2}\Bigg).$$

На следующем этапе преобразования разложим правую часть равенства на множители следующим образом:

$$\sum_{i=1}^{n}x_{1,i}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)=$$

$$=n\cdot\left(\sum_{i=1}^{n}x_{1,i}^2\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}\right)-\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}\right)-\right.$$

$$\left.-\left(\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}-\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}\right)^2\right)=$$

$$=n\cdot\left(\left(\sum_{i=1}^{n}x_{1,i}^2-\frac{\left(\sum_{i=1}^{n}x_{1,i}\right)^2}{n}\right)\cdot\left(\sum_{i=1}^{n}x_{2,i}^2-\frac{\left(\sum_{i=1}^{n}x_{2,i}\right)^2}{n}\right)-\right.$$

$$\left.-\left(\sum_{i=1}^{n}x_{1,i}\cdot x_{2,i}-\frac{\sum_{i=1}^{n}x_{1,i}\cdot\sum_{i=1}^{n}x_{2,i}}{n}\right)^2\right).$$

Далее, подставляя (155), получим

$$\sum_{i=1}^{n}x_{1,i}\cdot\left(B_3\cdot B_4-B_2\cdot B_5\right)=n\cdot\left(C_2\cdot C_4-C_3^2\right). \qquad (180)$$

Теперь с учетом (171) и (173) преобразуем выражение, находящееся в скобках в числителе (179):

$$B_5 \cdot \sum_{i=1}^{n} x_{1,i}^2 - B_3 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} = \left(\sum_{i=1}^{n} x_{2,i}^2 - \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) \cdot \sum_{i=1}^{n} x_{1,i}^2 -$$

$$- \left(\sum_{i=1}^{n} x_{2,i} \cdot x_{1,i} - \frac{\sum_{i=1}^{n} x_{1,i}^2 \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \right) \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \cdot$$

Раскроем скобки в правой части полученного равенства

$$B_5 \cdot \sum_{i=1}^{n} x_{1,i}^2 - B_3 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} = \sum_{i=1}^{n} x_{1,i}^2 \cdot \sum_{i=1}^{n} x_{2,i}^2 -$$

$$- \frac{\sum_{i=1}^{n} x_{2,i} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 +$$

$$+ \frac{\sum_{i=1}^{n} x_{1,i}^2 \cdot \sum_{i=1}^{n} x_{2,i}}{\sum_{i=1}^{n} x_{1,i}} \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i}$$

и приведем подобные члены:

$$B_5 \cdot \sum_{i=1}^{n} x_{1,i}^2 - B_3 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} = \sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 . \tag{181}$$

Теперь с учетом (180) и (181) преобразуем выражение (179) следующим образом:

$$V(a_0) = \frac{\sigma^2 \cdot \left(B_5 \cdot \sum_{i=1}^{n} x_{1,i}^2 - B_3 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)}{\sum_{i=1}^{n} x_{1,i} \cdot (B_3 \cdot B_4 - B_2 \cdot B_5)} =$$

$$= \frac{\sigma^2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 \right)}{n \cdot \left(C_2 \cdot C_4 - C_3^2 \right)} .$$

В том случае, когда σ^2 неизвестна, предполагая, что модель корректна, вместо нее можно использовать оценку s^2 – остаточный средний квадрат. Тогда дисперсию коэффициента регрессии a_0 можно оценить следующим образом:

$$s_{a_0}^2 = \frac{s^2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 \right)}{n \cdot \left(C_2 \cdot C_4 - C_3^2 \right)}. \qquad (182)$$

Теперь, зная дисперсию коэффициента a_0 (182), вычислим t-критерий для проверки гипотезы о равенстве этого коэффициента заданному числу $a_{0,0}$.

Воспользовавшись формулой (144) и учитывая, что стандартное отклонение есть квадратный корень из дисперсии, t-критерий для a_0 можно вычислить следующим образом:

$$t = (a_0 - a_{0,0}) \cdot \left(\frac{n \cdot \left(C_2 \cdot C_4 - C_3^2 \right)}{s^2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 \right)} \right)^{1/2}$$

или

$$t = a_0 \cdot \left(\frac{n \cdot \left(C_2 \cdot C_4 - C_3^2 \right)}{s^2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 \right)} \right)^{1/2} \qquad (183)$$

для проверки гипотезы о равенстве коэффициента регрессии a_0 нулю.

Мы рассмотрели варианты с одной и двумя независимыми переменными. В общем случае, когда имеется много независимых переменных, дисперсия коэффициента a_j равна соответствующему диагональному элементу ковариационной матрицы вектора коэффициентов регрессии a_j. Эта матрица представляет собой результат обращения матрицы коэффициентов при параметрах a_j в системе нормальных уравнений с последующим умножением на дисперсию зависимой переменной относительно линии регрессии $V(y_i) = \sigma^2$.

7.3. Оценка коэффициентов прямой

Рассмотрим процесс построения самой простой регрессионной модели в случае, когда у нас есть одна независимая переменная. Такая модель является частным случаем линейной регрессионной модели (123) и имеет вид прямой:

$$y = a_0 + a_1 \cdot x.$$

Для того чтобы получить оценку коэффициентов a_0 и a_1 уравнения прямой линии методом наименьших квадратов, следует решить систему нормальных уравнений. В случае одной независимой переменной система (131) примет следующий вид:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ a_0 \cdot \sum_{i=1}^{n} x_i + a_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \cdot x_i. \end{cases}$$

Вычислим коэффициенты регрессии прямой линии по данным замеров диаметров и высот деревьев в древостое. Для этого воспользуемся результатами группировки, приведенными в таблице распределения 14. В случае использования сгруппированных исходных данных некоторые пары значений зависимой и независимой переменной могут встречаться несколько раз. В связи с этим система нормальных уравнений примет следующий вид:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{k} f_i \cdot x_i = \sum_{i=1}^{k} f_i \cdot y_i; \\ a_0 \cdot \sum_{i=1}^{k} f_i \cdot x_i + a_1 \cdot \sum_{i=1}^{k} f_i \cdot x_i^2 = \sum_{i=1}^{k} f_i \cdot y_i \cdot x_i. \end{cases} \qquad (184)$$

Для того чтобы вычислить все необходимые суммы, составим вспомогательную таблицу (табл. 44).

В данной таблице суммы вычисляются сначала по интервалам, а затем складываются. Подставив значения сумм в систему нормальных уравнений (184), получим

$$\begin{cases} a_0 \cdot 200 + a_1 \cdot 6320{,}40 = 4960{,}4; \\ a_0 \cdot 6320{,}40 + a_1 \cdot 210797{,}2 = 158\,989{,}1. \end{cases} \qquad (185)$$

Таблица 44. **Вспомогательная таблица для вычисления коэффициентов регрессии прямой**

H \ D	17,65	20,85	24,05	27,25	30,45	33,65	36,85	40,05	43,25	46,45	49,65	52,85	56,05	Сумма	$\sum f_j \cdot y_j$
30,05	–	–	–	1				–	–	–	–			1	30,1
29,05	–	–	–	1	1			–	–	–	1			3	87,2
28,05	–	–	1	1	2	1	3	2	–	1	2	1		14	392,7
27,05	–	–	–	1	3	4	5	3	4	2	–		2	24	649,2
26,05	–	–	1	7	4	13	10	3	4	–	1			43	1120,2
25,05	–	–	1	7	7	12	2	1	1	–	–	1		32	801,6
24,05	–	–	7	7	12	1	3	1	–	–	–			31	745,6
23,05	–	1	8	12	2	2		–	–	–	–			25	576,3
22,05	–	3	8	1	1			–	–	–	–			13	286,7
21,05	–	1	–	1				–	–	–	–			2	42,1
20,05	–	4	1					–	–	–	–			5	100,3
19,05	1	2	1					–	–	–	–			4	76,2
18,05	–	–	1					–	–	–	–			1	18,1
17,05	2	–	–					–	–	–	–			2	34,1
f_x	3	11	29	39	32	33	23	10	9	3	4	2	2	200	4960,4
$\sum f_i \cdot x_i$	52,95	229,35	697,45	1062,75	974,40	1110,45	847,55	400,50	389,25	139,35	198,60	105,70	112,10	6320,40	–
$\sum f_i \cdot x_i^2$	934,6	4781,9	16773,7	28959,9	29670,5	37366,6	31232,2	16040,0	16835,1	6472,8	9860,5	5586,2	6283,2	210797,2	–
$\sum f_{i,j} \cdot y_j \cdot x_i$	938,1	4765,3	16004,1	26158,6	24439,2	28456,1	22189,2	10593,2	10269,7	3815,9	5521,1	2806,3	3032,3	158989,1	–
$\widetilde{y}_i$	22,00	22,64	23,28	23,93	24,57	25,21	25,86	26,5	27,14	27,79	28,43	29,07	29,72	–	–
$\sum f_{i,j} \cdot (y_j - \widetilde{y}_i)^2$	57,71	56,35	105,91	151,84	87,33	41,76	32,97	14,43	9,15	1,16	6,34	17,2	14,26	596,41	–

Решим полученную систему уравнений. Для этого разделим каждое из уравнений системы (185) на коэффициенты при параметре a_0:

$$\begin{cases} a_0 + a_1 \cdot 31{,}602 = 24{,}802; \\ a_0 + a_1 \cdot 33{,}352 = 25{,}155. \end{cases} \tag{186}$$

Теперь вычтем первое уравнение системы (186) из второго

$$a_1 \cdot 1{,}750 = 0{,}353 \tag{187}$$

и выразим из полученного уравнения (187) коэффициент a_1:

$$a_1 = \frac{0{,}353}{1{,}750} = 0{,}2017\,.$$

Подставляя вычисленное значение коэффициента a_1 в первое уравнение системы (186) и выразив из него коэффициент a_0, получим

$$a_0 = 24{,}802 - a_1 \cdot 31{,}602 = 24{,}802 - 0{,}2017 \cdot 31{,}602 = 18{,}43\,.$$

Таким образом, у нас получилась регрессионная модель зависимости высоты от диаметра деревьев в сосновом древостое следующего вида:

$$\tilde{y} = 18{,}43 + 0{,}2017 \cdot x\,, \tag{188}$$

или, используя другие обозначения

$$\tilde{h} = 18{,}43 + 0{,}2017 \cdot d\,.$$

На рис. 28 изображено полученное регрессионное уравнение прямой линии.

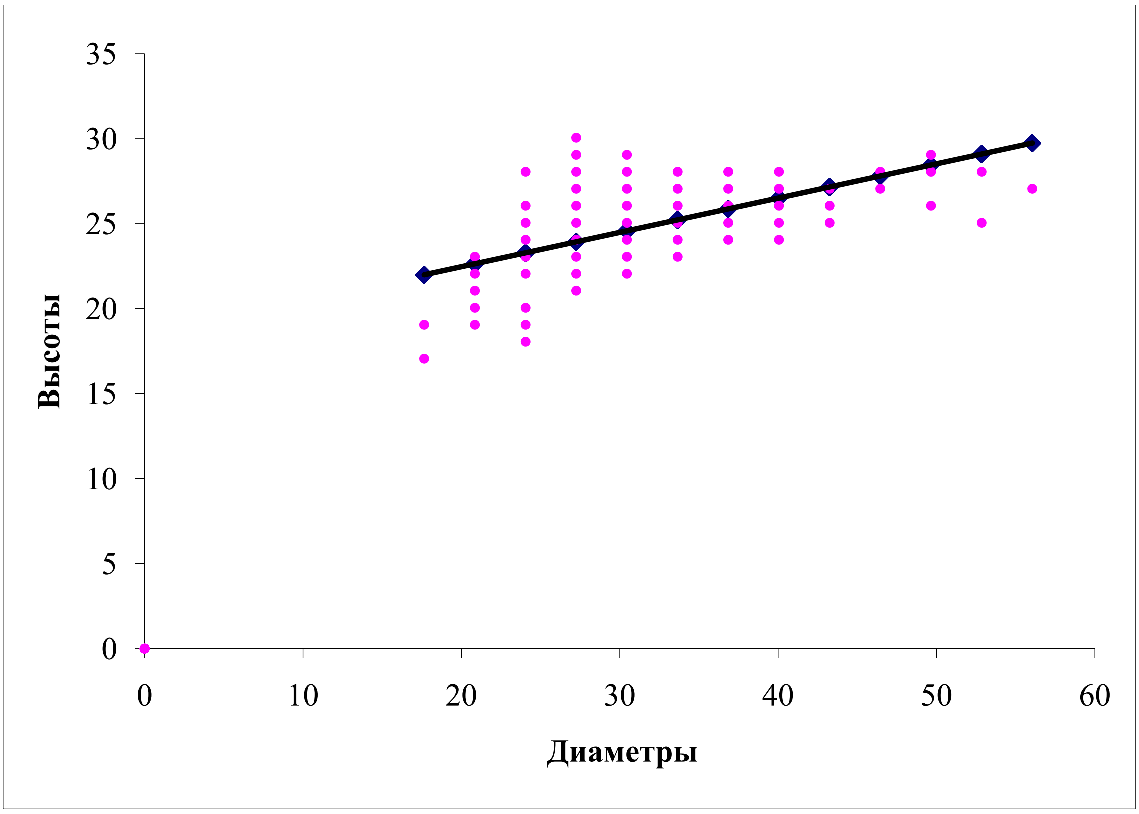


Рис. 28. Зависимость между высотами и диаметрами деревьев в древостое (прямая)

Проанализируем полученное уравнение. Для начала рассчитаем таблицу дисперсионного анализа (табл. 45). Число степеней свободы для суммы квадратов, обусловленной регрессией, в случае уравнения прямой линии равно 1, так как, помимо свободного члена, в нем есть только один коэффициент регрессии a_1. Общей скорректированной сумме квадратов соответствует число степеней свободы, на единицу меньшее общего количества наблюдений, т. е. 200–1 = 199. На сумму квадратов отклонений относительно регрессии остается 199–1 = 198 степеней свободы. Пользуясь полученным регрессионным уравнением прямой линии (188), определим теоретические высоты $\tilde{y}_i$ и сумму квадратов отклонений эмпирических высот от теоретических (табл. 44). Это значение 596,41 является суммой квадратов отклонений относительно регрессии, или остатком (табл. 45). Общая скорректированная сумма квадратов отклонений – это сумма квадратов отклонений всех высот от средней арифметической высоты. Данная величина была получена ранее для рассматриваемых данных при вычислении показателей вариации (табл. 23). Сумму квадратов отклонений, обусловленную регрессией, получим как разницу между общей скорректированной суммой квадратов отклонений и остатком: $1051{,}51 - 596{,}52 = 454{,}99$. Разделив суммы квадратов на соответствующие числа степеней свободы, получим средний квадрат, обусловленный регрессией (454,99), и средний квадрат относительно регрессии (3,01).

Таблица 45. **Дисперсионный анализ уравнения прямой линии**

Источник вариации	Число степеней свободы	Суммы квадратов отклонений *SS*	Средние квадраты *MS*
Обусловленный регрессией	1	454,99	454,99
Относительно регрессии (остаток)	198	596,52	3,01
Общий, скорректированный	199	1051,51	

Теперь рассчитаем F-статистику Фишера, воспользовавшись формулой (141):

$$F = \frac{MS_R}{s^2} = \frac{454{,}99}{3{,}01} = 151{,}16.$$

С помощью полученного значения проверим гипотезу об отсутствии связи между зависимой и независимой переменными. Для этого в табл. 8 приложения найдем квантиль распределения Фишера с 1 и 198 степенями свободы для уровня значимости $\alpha = 0{,}05$. Это значение равно $F_{0{,}05,\,1,\,198} = 3{,}84$. Так как вычисленное значение F-критерия Фишера превышает табличное, мы отвергаем нулевую гипотезу об отсутствии связи между высотами и диаметрами. Следовательно, результаты расчетов подтверждают наличие линейной связи между диаметрами и высотами деревьев в древостое.

Наряду с дисперсионным анализом вычислим коэффициент детерминации, пользуясь формулой (142)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{454{,}99}{1051{,}51} = 0{,}4327;$$

приведенную R^2 статистику с помощью выражения (143)

$$R_a^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1} =$$

$$= 1 - (1 - 0{,}4327) \cdot \frac{200-1}{200-1-1} = 0{,}4298$$

и коэффициент корреляции, который является квадратным корнем из коэффициента детерминации:

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{\frac{454{,}99}{1051{,}51}} = 0{,}6578.$$

Полученные значения последних показателей позволяют оценить степень адекватности линейной модели связи между высотами и диаметрами экспериментальным данным как среднюю.

Следующий этап оценки качества полученного уравнения регрессии будет заключаться в проверке гипотез о равенстве коэффициентов регрессии нулю. Для этого следует вычислить t-статистики Стьюдента для коэффициентов a_0 и a_1 с помощью формул (152) и (147). Необходимые для вычислений суммы мы можем найти в табл. 20, 22. Остаточный средний квадрат s^2 возьмем из табл. 45:

$$t_{a_0} = \frac{a_0 \cdot \left(n \cdot \sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2\right)^{1/2}}{\left(s^2 \cdot \sum_{i=1}^{n} x_{1,i}^2\right)^{1/2}} = \frac{18{,}43 \cdot (200 \cdot 11\,060{,}04)^{1/2}}{(3{,}01 \cdot 210\,797{,}2)^{1/2}} =$$

$$= \frac{27\,410{,}609}{796{,}55} = 34{,}41;$$

$$t_{a_1} = a_1 \cdot \left(\frac{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2}{s^2} \right)^{1/2} = 0{,}2017 \cdot \left(\frac{11\,060{,}04}{3{,}01} \right)^{1/2} =$$

$$= 0{,}2017 \cdot 60{,}62 = 12{,}23.$$

Далее, для проверки гипотезы о равенстве коэффициентов регрессии a_0 и a_1 нулю в табл. 5 приложения следует найти значение квантиля t-распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и 198 степеней свободы $t_{0{,}05;198} = 1{,}653$. Так как абсолютные величины вычисленных t-статистик больше табличных $\left|t_{a_0}\right| = |34{,}41| > 1{,}653$ и $\left|t_{a_1}\right| = |12{,}23| > 1{,}653$, то нулевую гипотезу о равенстве нулю коэффициентов регрессии следует отвергнуть. Следовательно, коэффициенты регрессии значимы.

Стандартная ошибка регрессионного уравнения прямой может быть вычислена по формуле (140), но в связи с тем, что исходные данные в рассматриваемом примере сгруппированы и вычисления ведутся на основе корреляционной решетки (табл. 14), удобнее воспользоваться следующим вариантом формулы:

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}}. \qquad (189)$$

Подставляя в формулу (189) необходимые значения из табл. 44 и учитывая, что в рассматриваемом примере $m = 1$, получим

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} = \sqrt{\frac{596{,}52}{200 - 1 - 1}} = \sqrt{3{,}01} = 1{,}735.$$

Используя данные дисперсионного анализа уравнения регрессии, приведенные в табл. 45, стандартную ошибку оценки можно определить, вычислив квадратный корень из среднего квадрата относительно регрессии:

$$s = \sqrt{3{,}01} = 1{,}735.$$

7.4. Оценка коэффициентов параболы второго порядка

Оценку коэффициентов параболы второго порядка методом наименьших квадратов дает решение системы нормальных уравнений:

$$
\begin{cases}
a_0 \cdot n + a_1 \cdot \sum_{i=1}^{k} f_i \cdot x_i + a_2 \cdot \sum_{i=1}^{k} f_i \cdot x_i^2 = \sum_{i=1}^{k} f_i \cdot y_i; \\
a_0 \cdot \sum_{i=1}^{k} f_i \cdot x_i + a_1 \cdot \sum_{i=1}^{k} f_i \cdot x_i^2 + a_2 \cdot \sum_{i=1}^{k} f_i \cdot x_i^3 = \sum_{i=1}^{k} f_i \cdot y_i \cdot x_i; \\
a_0 \cdot \sum_{i=1}^{k} f_i \cdot x_i^2 + a_1 \cdot \sum_{i=1}^{k} f_i \cdot x_i^3 + a_2 \cdot \sum_{i=1}^{k} f_i \cdot x_i^4 = \sum_{i=1}^{k} f_i \cdot y_i \cdot x_i^2.
\end{cases} \qquad (190)
$$

Вычислим коэффициенты уравнения параболы второго порядка, описывающего связь высот и диаметров деревьев в древостое. Для выполнения вычислений составим вспомогательную таблицу, аналогичную той, которую составляли для регрессионного уравнения прямой (табл. 46). Подставив найденные значения в систему нормальных уравнений (190), получим

$$
\begin{cases}
a_0 \cdot 200 + a_1 \cdot 6320{,}40 + a_2 \cdot 210797{,}2 = 4960{,}4; \\
a_0 \cdot 6320{,}40 + a_1 \cdot 210797{,}2 + a_2 \cdot 7428682{,}1 = 158989{,}1; \\
a_0 \cdot 210797{,}2 + a_1 \cdot 7428682{,}1 + a_2 \cdot 276643758{,}2 = 5371094{,}7.
\end{cases} \qquad (191)
$$

Таблица 46. **Вспомогательная таблица для вычисления коэффициентов регрессии параболы**

H \ D	17,65	20,85	24,05	27,25	30,45	33,65	36,85	40,05	43,25	46,45	49,65	52,85	56,05	Сумма	$\sum f_i \cdot y_i$
30,05	–	–	–	1	–	–	–	–	–	–	–	–	–	1	30,1
29,05	–	–	–	1	1	–	–	–	–	–	1	–	–	3	87,2
28,05	–	–	1	1	2	1	3	2	–	1	2	1	–	14	392,7
27,05	–	–	–	1	3	4	5	3	4	2	–	–	2	24	649,2
26,05	–	–	1	7	4	13	10	3	4	–	1	–	–	43	1120,2
25,05	–	–	1	7	7	12	2	1	1	–	–	1	–	32	801,6
24,05	–	–	7	7	12	1	3	1	–	–	–	–	–	31	745,6
23,05	–	1	8	12	2	2	–	–	–	–	–	–	–	25	576,3
22,05	–	3	8	1	1	–	–	–	–	–	–	–	–	13	286,7
21,05	–	1	0	1	–	–	–	–	–	–	–	–	–	2	42,1
20,05	–	4	1	–	–	–	–	–	–	–	–	–	–	5	100,3
19,05	1	2	1	–	–	–	–	–	–	–	–	–	–	4	76,2
18,05	–	–	1	–	–	–	–	–	–	–	–	–	–	1	18,1
17,05	2	–	–	–	–	–	–	–	–	–	–	–	–	2	34,1
f_x	3	11	29	39	32	33	23	10	9	3	4	2	2	200	4960,4
$\sum f_i \cdot x_i$	52,95	229,35	697,45	1062,75	974,40	1110,45	847,55	400,50	389,25	139,35	198,60	105,70	112,10	6320,40	
$\sum f_i \cdot x_i^2$	934,6	4781,9	16 773,7	28 959,9	29 670,5	37 366,6	31 232,2	16 040,0	16 835,1	6472,8	9860,5	5586,2	6283,2	210 797,2	
$\sum f_i \cdot x_i^3$	16 495,7	99 702,6	403 407,5	789 157,3	903 466,7	1 257 386,1	1 150 906,6	642 402,0	728 118,1	300 661,6	489 573,8	295 230,7	352 173,4	7 428 682,1	
$\sum f_i \cdot x_i^4$	291 138,8	2 078 820,2	9 701 934,1	21 504 563,6	27 510 543,2	42 311 090,1	42 410 930,9	25 728 240,2	31 491 036,6	13 965 745,6	24 307 315,8	15 603 066,6	19 739 332,5	276 643 758,2	
$\sum f_i \cdot y_i \cdot x_i$	938,1	4765,3	16 004,1	26 158,6	24 439,2	28 456,1	22 189,2	10 593,2	10 269,7	3815,9	5521,1	2806,3	3032,3	158 989,1	
$\sum f_i \cdot y_i \cdot x_i^2$	16 557,5	99 356,5	384 898,6	712 821,9	744 173,6	957 547,8	817 672,0	424 257,7	444 164,5	177 248,6	274 122,6	148 313,0	169 960,4	5 371 094,7	
$\widetilde{y}_i$	19,89	21,46	22,83	24,01	25,01	25,81	26,42	26,84	27,07	27,11	26,96	26,62	26,08		
$\sum f_{i,j} \cdot (y_j - \widetilde{y}_i)^2$	16,84	23,31	103,08	147,82	79,13	37,18	31,93	15,92	8,24	0,89	7,57	4,51	1,88	478,30	

Найдем решение системы нормальных уравнений (191). Сначала разделим все уравнения системы на коэффициенты при параметре a_0:

$$\begin{cases} a_0 + a_1 \cdot 31{,}6020 + a_2 \cdot 1053{,}9860 = 24{,}8020; \\ a_0 + a_1 \cdot 33{,}3519 + a_2 \cdot 1175{,}3500 = 25{,}1549; \\ a_0 + a_1 \cdot 35{,}2409 + a_2 \cdot 1312{,}3692 = 25{,}4799. \end{cases} \quad (192)$$

Теперь вычтем первое уравнение системы (192) из второго, а второе – из третьего. В результате получим систему из двух уравнений:

$$\begin{cases} a_1 \cdot 1{,}7499 + a_2 \cdot 121{,}3640 = 0{,}3529; \\ a_1 \cdot 1{,}8890 + a_2 \cdot 137{,}0192 = 0{,}3250. \end{cases} \quad (193)$$

Теперь вновь разделим уравнения системы (193) на коэффициент, на этот раз при параметре a_1:

$$\begin{cases} a_1 + a_2 \cdot 69{,}35\,482 = 0{,}20\,167; \\ a_1 + a_2 \cdot 72{,}53\,531 = 0{,}17\,205. \end{cases} \quad (194)$$

Вычитая первое уравнение системы (194) из второго, получим

$$a_2 \cdot 3{,}18\,049 = -0{,}02\,962,$$

откуда нетрудно выразить параметр a_2:

$$a_2 = \frac{-0{,}02\,962}{3{,}18\,049} = -0{,}009\,313.$$

Подставив полученное значение параметра a_2 в первое уравнение системы (194), выразим из него и вычислим величину параметра a_1:

$$\begin{aligned} a_1 &= 0{,}20\,167 - a_2 \cdot 69{,}35\,482 = \\ &= 0{,}20\,167 - (-0{,}009\,313 \cdot 69{,}35\,482) = 0{,}8476. \end{aligned} \quad (195)$$

Теперь, воспользовавшись первым уравнением из системы (192), а также значениями параметров a_1 и a_2, вычислим величину a_0:

$$\begin{aligned} a_0 &= 24{,}8020 - a_1 \cdot 31{,}6020 - a_2 \cdot 1053{,}9860 = \\ &= 24{,}8020 - 0{,}8476 \cdot 31{,}6020 - (-0{,}009\,313 \cdot 1053{,}9860) = \\ &= 24{,}8020 - 26{,}7859 + 9{,}8158 = 7{,}8319. \end{aligned} \quad (196)$$

В результате выполненных вычислений мы получили регрессионное уравнение параболы второго порядка, описывающее зависимость высоты от диаметра в чистом сосновом древостое:

$$\tilde{y} = 7{,}8319 + 0{,}8476 \cdot x - 0{,}009\,313 \cdot x^2, \qquad (197)$$

или, с использованием других обозначений

$$\tilde{h} = 7{,}8319 + 0{,}8476 \cdot d - 0{,}009\,313 \cdot d^2.$$

На рис. 29 изображено полученное регрессионное уравнение параболы второго порядка.

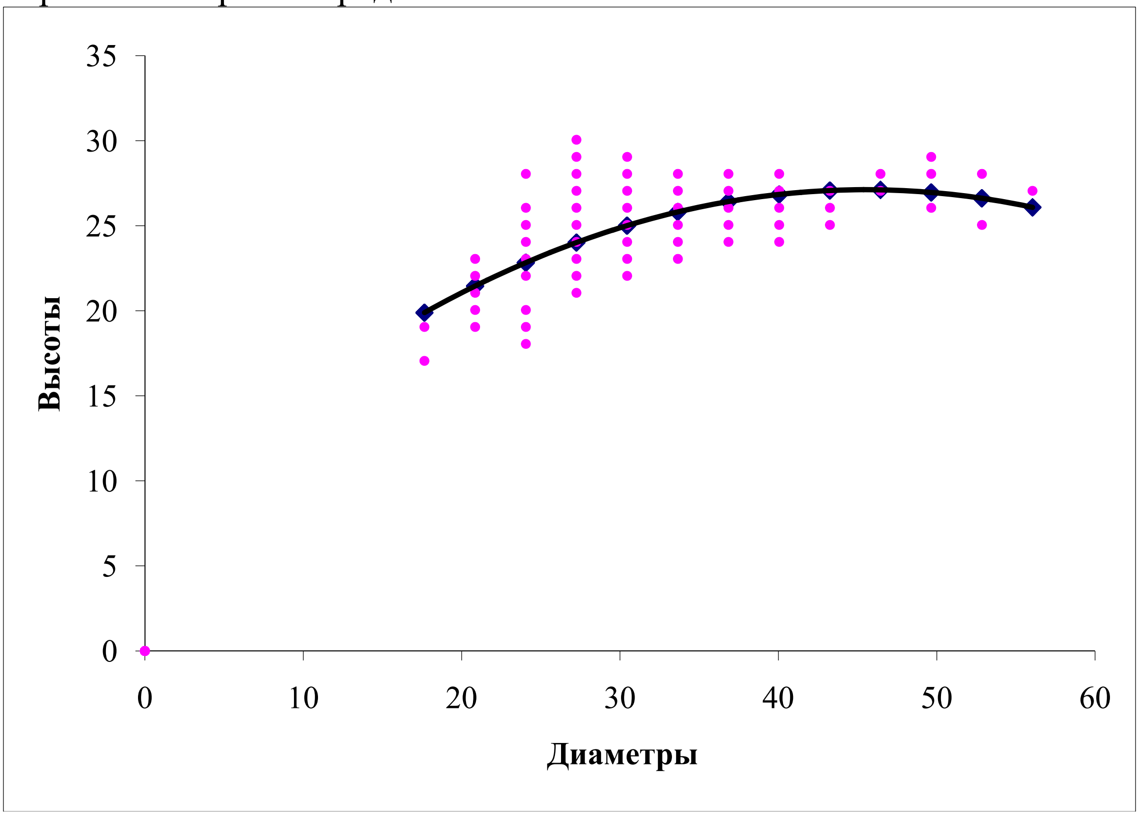


Рис. 29. Зависимость между высотами и диаметрами деревьев в древостое (парабола второго порядка)

Для анализа полученного уравнения рассчитаем таблицу дисперсионного анализа (табл. 47). Число степеней свободы для суммы квадратов, обусловленной регрессией, в случае квадратной параболы равно 2, так как, помимо свободного члена, в этом уравнении есть два коэффициента регрессии – a_1 и a_2. Общей скорректированной сумме квадратов соответствует число степеней свободы, на единицу меньшее общего количества наблюдений, т. е. $200 - 1 = 199$. На сумму квадратов отклонений относительно регрессии остается $199 - 2 = 197$ степеней свободы. Пользуясь полученным регрессионным уравнением параболы второго порядка (197), определим теоретические высоты $\tilde{y}_i$ и сумму квадратов отклонений эмпирических высот от теоретических (табл. 46). Это значение 478,30 является суммой квадратов отклонений относительно регрессии, или остатком (табл. 47).

Таблица 47. **Дисперсионный анализ уравнения параболы второго порядка**

Источник вариации	Число степеней свободы	Суммы квадратов отклонений *SS*	Средние квадраты *MS*
Обусловленный регрессией	2	573,21	286,61
Относительно регрессии (остаток)	197	478,30	2,43
Общий, скорректированный	199	1051,51	–

Общая скорректированная сумма квадратов отклонений – это сумма квадратов отклонений всех высот от средней арифметической высоты. Данная величина была получена ранее для рассматриваемых данных при вычислении показателей вариации (табл. 23), и мы ее уже использовали для расчета таблицы дисперсионного анализа при исследовании регрессионного уравнения прямой линии (табл. 45). Сумму квадратов отклонений, обусловленную регрессией, можно вычислить как разницу между общей скорректированной суммой квадратов отклонений и остатком: 1051,51 – 478,30 = 573,21. Разделив суммы квадратов на соответствующие числа степеней свободы, получим средний квадрат, обусловленный регрессией (286,61), и средний квадрат относительно регрессии (2,43).

Теперь мы можем вычислить F-статистику Фишера, применив формулу (141):

$$F = \frac{MS_R}{s^2} = \frac{286,61}{2,43} = 117,95\,.$$

Пользуясь полученным значением, проверим гипотезу, заключающуюся в предположении о равенстве нулю всех коэффициентов регрессии уравнения параболы второго порядка, за исключением свободного члена. Для этого в табл. 8 приложения найдем квантиль распределения Фишера с 2 и 197 степенями свободы для уровня значимости $\alpha = 0,05$. Это значение равно $F_{0,05,\,2,\,197} = 3,00$. Так как вычисленное значение F-критерия Фишера превышает табличное, мы отвергаем проверяемую нулевую гипотезу.

Следовательно, результаты расчетов подтверждают значимость полученного регрессионного уравнения параболы второго порядка, описывающего связь между диаметрами и высотами деревьев в древостое.

Далее, пользуясь формулой (142), вычислим коэффициент детерминации

$$R^2 = \frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{573,21}{1051,51} = 0,5451.$$

Приведенная R^2 статистика для уравнения параболы второго порядка, вычисленная с помощью выражения (143), будет равна:

$$R_a^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1} =$$

$$= 1 - (1 - 0,5451) \cdot \frac{200-1}{200-2-1} = 0,5405.$$

Теперь извлечем квадратный корень из коэффициента детерминации, для того чтобы получить коэффициент корреляции:

$$R = \sqrt{R^2} = \sqrt{0,5451} = 0,7383.$$

Величина коэффициента корреляции 0,7383 позволяет оценить степень адекватности параболической модели связи между высотами и диаметрами экспериментальным данным как высокую.

Следующий этап оценки качества полученного уравнения регрессии будет заключаться в проверке гипотез о равенстве коэффициентов регрессии нулю. Для этого следует вычислить t-статистики Стьюдента для коэффициентов a_0, a_1 и a_2. Для того чтобы можно было воспользоваться формулами (183), (168) и (162), необходимо вычислить значения C_2, C_3 и C_4 с помощью формул (155). В случае параболы второго порядка переменная $x_1 = x$, а переменная $x_2 = x^2$. Тогда уравнения (155) будут выглядеть следующим образом:

$$C_2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n};$$

$$C_3 = \sum_{i=1}^{n} x_i^3 - \frac{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} x_i}{n};$$

$$C_4 = \sum_{i=1}^{n} x_i^4 - \frac{\left(\sum_{i=1}^{n} x_i^2\right)^2}{n}.$$

Учитывая то, что вычисления выполняются на основе сгруппированных данных – корреляционной решетки, формулы для вычисления значений C_2, C_3 и C_4 представим следующим образом:

$$C_2 = \sum_{i=1}^{k} f_i \cdot x_i^2 - \frac{\left(\sum_{i=1}^{k} f_i \cdot x_i\right)^2}{n};$$

$$C_3 = \sum_{i=1}^{k} f_i \cdot x_i^3 - \frac{\sum_{i=1}^{k} f_i \cdot x_i^2 \cdot \sum_{i=1}^{k} f_i \cdot x_i}{n};$$

$$C_4 = \sum_{i=1}^{k} f_i \cdot x_i^4 - \frac{\left(\sum_{i=1}^{k} f_i \cdot x_i^2\right)^2}{n},$$

где k – количество классов (количество ячеек корреляционной решетки); x_i – значение класса вариационного ряда диаметров, соответствующее i-той ячейки корреляционной решетки; f_i – частота i-того класса (i-той ячейке корреляционной решетки).

Необходимые для вычислений суммы мы можем найти в табл. 20, 46. Остаточный средний квадрат s^2 возьмем из табл. 47:

$$C_2 = 210\,797{,}2 - \frac{(6320{,}40)^2}{200} = 11\,059{,}9;$$

$$C_3 = 7\,428\,682{,}1 - \frac{210\,797{,}2 \cdot 6320{,}40}{200} = 767\,069{,}0; \qquad (198)$$

$$C_4 = 276\,643\,758{,}2 - \frac{(210\,797{,}2)^2}{200} = 54\,466\,460{,}6.$$

Теперь, пользуясь формулами (183) (168) и (162), а также учитывая, что переменная $x_1 = x$, а переменная $x_2 = x^2$, вычислим t-критерии Стьюдента для проверки гипотез о равенстве коэффициентов регрессии a_0, a_1 и a_2 нулю. Для этого подставим вычисленные ранее значения C_2, C_3 и C_4 (198), величину коэффициента a_0 (196) и остаточный средний квадрат s^2 из табл. 47 в формулу (183) и получим t-критерий для коэффициента регрессии a_0:

$$t_{a_0} = a_0 \cdot \left(\frac{n \cdot \left(C_2 \cdot C_4 - C_3^2 \right)}{s^2 \cdot \left(\sum_{i=1}^{n} x_{2,i}^2 \cdot \sum_{i=1}^{n} x_{1,i}^2 - \left(\sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} \right)^2 \right)} \right)^{1/2} =$$

$$= 7{,}8319 \cdot \left(200 \cdot \left(11\,059{,}9 \cdot 54\,466\,460{,}6 - 767\,069{,}0^2\right)\right)^{1/2} \Big/$$

$$\Big/ \left(2{,}43 \cdot \left(276\,643\,758{,}2 \cdot 210\,797{,}2 - 7\,428\,682{,}1^2\right)\right)^{1/2} =$$

$$= 7{,}8319 \cdot 1\,673\,245{,}758 \Big/ \left(2.43 \cdot 3\,130\,411\,883\,176{,}63\right)^{1/2} =$$

$$= 7{,}8319 \cdot 1\,673\,245{,}758 / 2\,758\,061{,}072 = 4{,}7514\,.$$

Подставив вычисленные значения C_2, C_3 и C_4 (198), величину коэффициента a_1 (195) и остаточный средний квадрат s^2 из табл. 47 в формулу (168), вычислим t-критерий для коэффициента регрессии a_1:

$$t_{a_1} = \frac{a_1}{s_{a_1}} = a_1 \cdot \left(\frac{C_2 \cdot C_4 - C_3^2}{s^2 \cdot C_4} \right)^{1/2} =$$

$$= 0{,}8476 \cdot \left(\frac{11\,059{,}9 \cdot 54\,466\,460{,}6 - 767\,069{,}0^2}{2{,}43 \cdot 54\,466\,460{,}6} \right)^{1/2} =$$

$$= 0{,}8476 \cdot 105{,}7\,679\,389^{1/2} = 0{,}8476 \cdot 10{,}2844 = 8{,}717\,.$$

Аналогичным образом по формуле (162) вычислим t-критерий и для коэффициента регрессии a_2:

$$t_{a_2} = \frac{a_2}{s_{a_2}} = a_2 \cdot \left(\frac{C_2 \cdot C_4 - C_3^2}{s^2 \cdot C_2} \right)^{1/2} =$$

$$= -0{,}009\,313 \cdot \left(\frac{11\,059{,}9 \cdot 54\,466\,460{,}6 - 767\,069{,}0^2}{2{,}43 \cdot 11\,059{,}9} \right)^{1/2} =$$

$$= -0{,}009\,313 \cdot 520\,873{,}1796^{1/2} =$$
$$= -0{,}009\,313 \cdot 721{,}7\,154\,423 = -6{,}721.$$

Далее, для проверки гипотезы о равенстве коэффициентов регрессии a_0, a_1 и a_2 нулю в табл. 5 приложения следует найти значение квантиля t-распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и 197 степеней свободы $t_{0,05;197} = 1{,}653$. Так как абсолютные величины вычисленных t-статистик для всех трех коэффициентов регрессии больше табличных $\left|t_{a_0}\right| = |4{,}7514| > 1{,}653$, $\left|t_{a_1}\right| = |8{,}717| > 1{,}653$ и $\left|t_{a_2}\right| = |-6{,}721| > 1{,}653$, то нулевую гипотезу о равенстве коэффициентов регрессии нулю следует отвергнуть. Следовательно, коэффициенты регрессии значимы.

Стандартная ошибка регрессионного уравнения параболы второго порядка может быть вычислена по формуле (140), но в связи с тем, что исходные данные в рассматриваемом примере сгруппированы и вычисления ведутся на основе корреляционной решетки (табл. 14), так же как и для уравнения прямой линии, удобнее воспользоваться формулой (189). Подставляя в формулу (189) необходимые значения из табл. 46 и учитывая, что для квадратной параболы $m = 2$, получим

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} = \sqrt{\frac{478{,}30}{200 - 2 - 1}} = \sqrt{2{,}43} = 1{,}559.$$

Используя данные дисперсионного анализа уравнения регрессии, приведенные в табл. 47, стандартную ошибку оценки можно определить, вычислив квадратный корень из среднего квадрата относительно регрессии:

$$s = \sqrt{2{,}43} = 1{,}559.$$

7.5. Оценка коэффициентов гиперболы

Рассмотрим процесс построения регрессионной модели, имеющей вид гиперболы:

$$y = a_0 + a_1/x. \qquad (199)$$

Для того чтобы получить оценку коэффициентов a_0 и a_1 уравнения прямой линии методом наименьших квадратов, следует решить систему нормальных уравнений. В случае гиперболы вида

(199) эта система будет выглядеть следующим образом:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i} = \sum\limits_{i=1}^{n} y_i; \\ a_0 \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i} + a_1 \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i^2} = \sum\limits_{i=1}^{n} \dfrac{y_i}{x_i}. \end{cases} \qquad (200)$$

Вычислим коэффициенты регрессии гиперболы по данным замеров диаметров и высот деревьев в древостое. Для этого воспользуемся результатами группировки, приведенными в таблице распределения 14. В связи с тем, что данные, на основе которых будут выполняться вычисления, сгруппированы, запишем систему нормальных уравнений (200) следующим образом:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum\limits_{i=1}^{k} \dfrac{f_i}{x_i} = \sum\limits_{i=1}^{k} f_i \cdot y_i; \\ a_0 \cdot \sum\limits_{i=1}^{k} \dfrac{f_i}{x_i} + a_1 \cdot \sum\limits_{i=1}^{k} \dfrac{f_i}{x_i^2} \cdot = \sum\limits_{i=1}^{k} \dfrac{f_i \cdot y_i}{x_i}. \end{cases} \qquad (201)$$

Для того чтобы вычислить все необходимые суммы, составим вспомогательную таблицу (табл. 48).

В данной таблице суммы вычисляются сначала по интервалам, а затем складываются. Подставив значения сумм в систему нормальных уравнений (201), получим

$$\begin{cases} a_0 \cdot 200 + a_1 \cdot 6{,}66\,674 = 4960{,}4; \\ a_0 \cdot 6{,}66\,674 + a_1 \cdot 0{,}233\,597 = 162{,}773. \end{cases} \qquad (202)$$

Решим полученную систему уравнений. Для этого разделим каждое из уравнений системы (202) на коэффициенты при параметре a_0:

$$\begin{cases} a_0 + a_1 \cdot 0{,}033\,334 = 24{,}802; \\ a_0 + a_1 \cdot 0{,}035\,039 = 24{,}416. \end{cases} \qquad (203)$$

Таблица 48. **Вспомогательная таблица для вычисления коэффициентов регрессионного уравнения гиперболы**

H \ D	17,65	20,85	24,05	27,25	30,45	33,65	36,85	40,05	43,25	46,45	49,65	52,85	56,05	Сумма	$\sum f_j \cdot y_j$
30,05	–	–	–	1	–	–	–	–	–	–	–	–	–	1	30,1
29,05	–	–	–	1	1	–	–	–	–	–	1	–	–	3	87,2
28,05	–	–	1	1	2	1	3	2	–	1	2	1	–	14	392,7
27,05	–	–	–	1	3	4	5	3	4	2	–	–	2	24	649,2
26,05	–	–	1	7	4	13	10	3	4	–	1	–	–	43	1120,2
25,05	–	–	1	7	7	12	2	1	1	–	–	1	–	32	801,6
24,05	–	–	7	7	12	1	3	1	–	–	–	–	–	31	745,6
23,05	–	1	8	12	2	2	–	–	–	–	–	–	–	25	576,3
22,05	–	3	8	1	1	–	–	–	–	–	–	–	–	13	286,7
21,05	–	1	–	1	–	–	–	–	–	–	–	–	–	2	42,1
20,05	–	4	1	–	–	–	–	–	–	–	–	–	–	5	100,3
19,05	1	2	1	–	–	–	–	–	–	–	–	–	–	4	76,2
18,05	–	–	1	–	–	–	–	–	–	–	–	–	–	1	18,1
17,05	2	–		–	–	–	–	–	–	–	–	–	–	2	34,1
f_x	3	11	29	39	32	33	23	10	9	3	4	2	2	200	4960,4
$\sum f_i / x_i$	0,16 997	0,52 758	1,20 582	1,43 119	1,05 090	0,98 068	0,62 415	0,24 969	0,20 809	0,06 459	0,08 056	0,03 784	0,03 568	6,66 674	–
$\sum f_i / x_i^2$	0,009 630	0,025 303	0,050 138	0,052 521	0,034 512	0,029 144	0,016 938	0,006 234	0,004 811	0,001 390	0,001 623	0,000 716	0,000 637	0,233 597	–
$\sum (f_{i,j} \cdot y_j) / x_i$	3,011	10,962	27,669	35,228	26,358	25,131	16,341	6,604	5,490	1,769	2,240	1,005	0,965	162,773	–
$\widetilde{y}_i$	19,53	21,50	22,94	24,05	24,92	25,62	26,21	26,70	27,12	27,48	27,79	28,07	28,31	–	–
$\sum f_{i,j} \cdot (y_j - \widetilde{y}_i)^2$	12,53	23,93	102,69	146,00	79,80	36,06	30,63	15,03	8,88	0,69	4,75	9,12	3,18	473,29	–

Теперь вычтем первое уравнение системы (203) из второго

$$a_1 \cdot 0{,}001\,705 = -0{,}386 \qquad (204)$$

и выразим из полученного уравнения (204) коэффициент a_1:

$$a_1 = \frac{-0{,}386}{0{,}001\,705} = -226{,}3\,.$$

Подставляя вычисленное значение коэффициента a_1 в первое уравнение системы (203) и выразив из него коэффициент a_0, получим

$$a_0 = 24{,}802 - a_1 \cdot 0{,}033\,334 = 24{,}802 + 226{,}3 \cdot 0{,}033\,334 = 32{,}35\,. (205)$$

Таким образом, у нас получилась регрессионная модель зависимости высоты от диаметра деревьев в сосновом древостое следующего вида:

$$\tilde{y} = 32{,}35 - \frac{226{,}3}{x}, \qquad (206)$$

или, используя другие обозначения

$$\tilde{h} = 32{,}35 - \frac{226{,}3}{d}.$$

На рис. 30 изображено полученное регрессионное уравнение прямой линии.

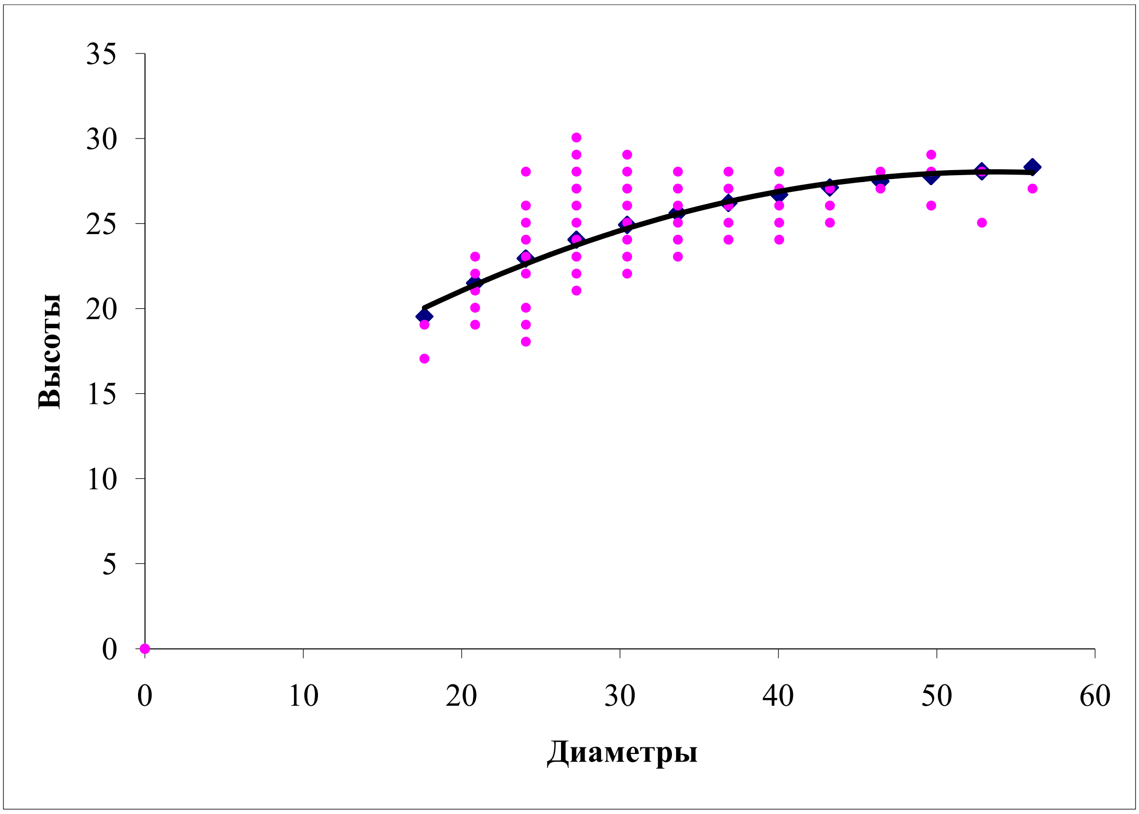


Рис. 30. Зависимость между высотами и диаметрами деревьев в древостое (гипербола)

Проанализируем полученное уравнение. Для начала рассчитаем таблицу дисперсионного анализа (табл. 49). Число степеней свободы для суммы квадратов, обусловленной регрессией, в случае уравнения гиперболы равно 1, так как, помимо свободного члена, в нем есть только один коэффициент регрессии – a_1. Общей скорректированной сумме квадратов соответствует число степеней свободы, на единицу меньшее общего количества наблюдений, т. е. 200 – 1 = 199. На сумму квадратов отклонений относительно регрессии остается 199 – 1 = 198 степеней свободы. Пользуясь полученным регрессионным уравнением гиперболы (206), определим теоретические высоты $\tilde{y}_i$ и сумму квадратов отклонений эмпирических высот от теоретических (табл. 48). Это значение 473,29 является суммой квадратов отклонений относительно регрессии, или остатком (табл. 49). Общая скорректированная сумма квадратов отклонений – это сумма квадратов отклонений всех высот от средней арифметической высоты. Данная величина была получена ранее для рассматриваемых данных при вычислении показателей вариации (табл. 23). Сумму квадратов отклонений, обусловленную регрессией, получим как разницу между общей скорректированной суммой квадратов отклонений и остатком: $1051{,}51 - 473{,}29 = 578{,}22$. Разделив суммы квадратов на соответствующие числа степеней свободы, получим средний квадрат, обусловленный регрессией (578,22), и средний квадрат относительно регрессии (2,39).

Таблица 49. **Дисперсионный анализ уравнения гиперболы**

Источник вариации	Число степеней свободы	Суммы квадратов отклонений *SS*	Средние квадраты *MS*
Обусловленный регрессией	1	578,22	578,22
Относительно регрессии (остаток)	198	473,29	2,39
Общий, скорректированный	199	1051,51	–

Теперь рассчитаем *F*-статистику Фишера, воспользовавшись формулой (141):

$$F = \frac{MS_R}{s^2} = \frac{578{,}22}{2{,}39} = 241{,}93\,.$$

С помощью полученного значения проверим гипотезу об отсутствии связи между зависимой и независимой переменными. Для этого в табл. 8 приложения найдем квантиль распределения Фишера с 1 и 198 степенями свободы для уровня значимости $\alpha = 0{,}05$. Это значение равно $F_{0{,}05,\,1,\,198} = 3{,}84$. Так как вычисленное значение *F*-критерия Фишера превышает табличное, мы отвергаем нулевую гипотезу об отсутствии связи между высотами и диаметрами. Следовательно, результаты расчетов подтверждают наличие гиперболической связи между диаметрами и высотами деревьев в древостое.

В дополнение к проведенному дисперсионному анализу вычислим коэффициент детерминации, пользуясь формулой (142)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{578{,}22}{1051{,}51} = 0{,}5499;$$

приведенную R^2 статистику с помощью выражения (143)

$$R_a^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1} =$$

$$= 1 - (1 - 0{,}5499) \cdot \frac{200-1}{200-1-1} = 0{,}5476$$

и коэффициент корреляции, который является квадратным корнем из коэффициента детерминации:

$$R = \sqrt{\frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \sqrt{\frac{578{,}22}{1051{,}51}} = 0{,}7415.$$

Полученные значения последних показателей позволяют оценить степень адекватности гиперболической модели связи между высотами и диаметрами экспериментальным данным как высокую.

Следующий этап оценки качества полученного уравнения регрессии будет заключаться в проверке гипотез о равенстве коэффициентов регрессии нулю. Для этого следует вычислить *t*-статистики Стьюдента для коэффициентов a_0 и a_1 с помощью формул (152) и (147). Учитывая (46), сумму квадратов отклонений значений x_i от их средней арифметической величины из числителя формулы (152) можно выразить следующим образом:

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = n \cdot S_x^2. \tag{207}$$

В свою очередь, дисперсию величины x, учитывая (53) и (18), можно представить в виде выражения

$$S^2 = \overline{x^2} - \left(\bar{x}\right)^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2. \tag{208}$$

Учитывая то, что в случае гиперболы $x_1 = 1/x$, формулу (152) можем представить в виде

$$t_{a_0} = \frac{a_0 \cdot \left(n \cdot \sum_{i=1}^{n}(1/x_i - \overline{1/x})^2\right)^{1/2}}{\left(s^2 \cdot \sum_{i=1}^{n} 1/x_i^2\right)^{1/2}}, \tag{209}$$

а с учетом (207) и (208) сумму квадратов отклонений из числителя полученной формулы преобразуем к выражению

$$\sum_{i=1}^{n}(1/x_i-\overline{1/x})^2=n\cdot\left(\frac{\sum_{i=1}^{n}1/x_i^2}{n}-\left(\frac{\sum_{i=1}^{n}1/x_i}{n}\right)^2\right)=$$

$$=\sum_{i=1}^{n}1/x_i^2-\frac{\left(\sum_{i=1}^{n}1/x_i\right)^2}{n}. \qquad (210)$$

Теперь, подставляя необходимые суммы из вспомогательной табл. 48 в выражение (210), вычислим значение суммы квадратов отклонений:

$$\sum_{i=1}^{n}(1/x_i-\overline{1/x})^2=0{,}233\,597-\frac{(6{,}66\,674)^2}{200}=$$

$$=0{,}233\,597-0{,}222\,227=0{,}01137. \qquad (211)$$

Далее, подставив полученное значение, остаточный средний квадрат s^2 из табл. 49 и сумму из табл. 48 в формулу (209), вычислим критерий Стьюдента для коэффициента регрессии a_0:

$$t_{a_0}=\frac{a_0\cdot\left(n\cdot\sum_{i=1}^{n}(1/x_i-\overline{1/x})^2\right)^{1/2}}{\left(s^2\cdot\sum_{i=1}^{n}1/x_i^2\right)^{1/2}}=$$

$$=\frac{32{,}35\cdot(200\cdot 0{,}01137)^{1/2}}{(2{,}39\cdot 0{,}233\,597)^{1/2}}=\frac{48{,}7831}{0{,}747\,193}=65{,}29.$$

Аналогичным образом, с учетом того, что в случае гиперболы $x_1 = 1/x$, преобразуем формулу (147), как показано ниже:

$$t_{a_1}=a_1\cdot\left(\frac{\sum_{i=1}^{n}(x_{1,i}-\bar{x}_1)^2}{s^2}\right)^{1/2}=a_1\cdot\left(\frac{\sum_{i=1}^{n}(1/x_i-\overline{1/x})^2}{s^2}\right)^{1/2}. \qquad (212)$$

Теперь, подставив в выражение (212) вычисленную ранее сумму квадратов отклонений (211) и остальные значения, определим t-статистику Стьюдента для коэффициента a_1:

$$t_{a_1} = a_1 \cdot \left(\frac{\sum_{i=1}^{n} (1/x_i - \overline{1/x})^2}{s^2} \right)^{1/2} = -226{,}3 \cdot \left(\frac{0{,}01137}{2{,}39} \right)^{1/2} = -15{,}61 .$$

Далее, для проверки гипотезы о равенстве коэффициентов регрессии a_0 и a_1 нулю, в табл. 5 приложения следует найти значение квантиля t-распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и 198 степеней свободы $t_{0,05;198} = 1{,}653$. Так как абсолютные значения вычисленных t-статистик больше табличных $\left| t_{a_0} \right| = \left| 65{,}29 \right| > 1{,}653$ и $\left| t_{a_1} \right| = \left| -15{,}61 \right| > 1{,}653$ то нулевую гипотезу о равенстве нулю коэффициентов регрессии следует отвергнуть. Следовательно, коэффициенты регрессии значимы.

Стандартная ошибка регрессионного уравнения гиперболы может быть вычислена по формуле (140), но в связи с тем, что исходные данные в рассматриваемом примере сгруппированы и вычисления ведутся на основе корреляционной решетки (табл. 14), удобнее воспользоваться следующим вариантом формулы:

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} .$$

Подставляя в формулу (189) необходимые значения из табл. 48 и учитывая, что в рассматриваемом примере $m = 1$, получим

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} = \sqrt{\frac{473{,}29}{200 - 1 - 1}} = 1{,}546 .$$

Используя данные дисперсионного анализа уравнения регрессии, приведенные в табл. 49, стандартную ошибку оценки можно определить, вычислив квадратный корень из среднего квадрата относительно регрессии:

$$s = \sqrt{2{,}39} = 1{,}546 .$$

7.6. Оценка коэффициентов логарифмической кривой типа $y = a_0 + a_1 \lg(x)$

Рассмотрим процесс построения логарифмической регрессионной модели, имеющей вид

$$y = a_0 + a_1 \cdot \lg(x). \tag{213}$$

Для того чтобы получить оценку коэффициентов регрессии a_0 и a_1 методом наименьших квадратов, следует решить систему нормальных уравнений. В случае логарифмической кривой вида (213) эта система будет выглядеть следующим образом:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum\limits_{i=1}^{n} \lg(x_i) = \sum\limits_{i=1}^{n} y_i; \\ a_0 \cdot \sum\limits_{i=1}^{n} \lg(x_i) + a_1 \cdot \sum\limits_{i=1}^{n} (\lg(x_i))^2 = \sum\limits_{i=1}^{n} y_i \cdot \lg(x_i). \end{cases} \tag{214}$$

Вычислим коэффициенты регрессии логарифмической кривой по данным замеров диаметров и высот деревьев в древостое. Для этого воспользуемся результатами группировки, приведенными в таблице распределения 14. В связи с тем, что данные, на основе которых будут выполняться вычисления, сгруппированы, запишем систему нормальных уравнений (214) следующим образом:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum\limits_{i=1}^{k} f_i \cdot \lg(x_i) = \sum\limits_{i=1}^{k} f_i \cdot y_i; \\ a_0 \cdot \sum\limits_{i=1}^{k} f_i \cdot \lg(x_i) + a_1 \cdot \sum\limits_{i=1}^{k} f_i \cdot (\lg(x_i))^2 = \sum\limits_{i=1}^{k} f_i \cdot y_i \cdot \lg(x_i). \end{cases} \tag{215}$$

Для того чтобы вычислить все необходимые суммы, составим вспомогательную таблицу (табл. 50).

В данной таблице суммы вычисляются сначала по интервалам, а затем складываются. Подставив значения сумм в систему нормальных уравнений (215), получим

$$\begin{cases} a_0 \cdot 200 + a_1 \cdot 297{,}652 = 4960{,}4; \\ a_0 \cdot 297{,}652 + a_1 \cdot 444{,}952 = 7413{,}90. \end{cases} \tag{216}$$

Решим полученную систему уравнений. Для этого разделим каждое из уравнений системы (216) на коэффициенты при параметре b_0:

$$\begin{cases} a_0 + a_1 \cdot 1{,}4883 = 24{,}802; \\ a_0 + a_1 \cdot 1{,}4949 = 24{,}908. \end{cases} \tag{217}$$

Теперь вычтем первое уравнение системы (217) из второго

$$a_1 \cdot 0{,}0066 = 0{,}106 \tag{218}$$

и выразим из полученного уравнения (218) коэффициент a_1:

$$a_1 = \frac{0{,}106}{0{,}0066} = 16{,}06.$$

Подставляя вычисленное значение коэффициента a_1 в первое уравнение системы (217) и выразив из него коэффициент a_0, имеем

$$a_0 = 24{,}802 - a_1 \cdot 1{,}4883 = 24{,}802 - 16{,}06 \cdot 1{,}4883 = 0{,}8999.$$

Таким образом, у нас получилась регрессионная модель зависимости высоты от диаметра деревьев в сосновом древостое следующего вида:

$$\tilde{y} = 0{,}8999 + 16{,}06 \cdot \lg(x), \tag{219}$$

или, используя другие обозначения

$$\tilde{h} = 0{,}8999 + 16{,}06 \cdot \lg(x).$$

Таблица 50. **Вспомогательная таблица для вычисления коэффициентов регрессионного уравнения логарифмической кривой**

H \ D	17,65	20,85	24,05	27,25	30,45	33,65	36,85	40,05	43,25	46,45	49,65	52,85	56,05	Сумма	$\sum f_j \cdot y_j$
30,05	–	–	–	1	–	–	–	–	–	–	–	–	–	1	30,1
29,05	–	–	–	1	1	–	–	–	–	–	1	–	–	3	87,2
28,05	–	–	1	1	2	1	3	2	–	1	2	1	–	14	392,7
27,05	–	–	–	1	3	4	5	3	4	2	–	–	2	24	649,2
26,05	–	–	1	7	4	13	10	3	4	–	1	–	–	43	1120,2
25,05	–	–	1	7	7	12	2	1	1	–	–	1	–	32	801,6
24,05	–	–	7	7	12	1	3	1	–	–	–	–	–	31	745,6
23,05	–	1	8	12	2	2	–	–	–	–	–	–	–	25	576,3
22,05	–	3	8	1	1	–	–	–	–	–	–	–	–	13	286,7
21,05	–	1	–	1	–	–	–	–	–	–	–	–	–	2	42,1
20,05	–	4	1	–	–	–	–	–	–	–	–	–	–	5	100,3
19,05	1	2	1	–	–	–	–	–	–	–	–	–	–	4	76,2
18,05	–	–	1	–	–	–	–	–	–	–	–	–	–	1	18,1
17,05	2	–	–	–	–	–	–	–	–	–	–	–	–	2	34,1
f_x	3	11	29	39	32	33	23	10	9	3	4	2	2	200	4960,4
$\sum f_i \cdot \lg(x_i)$	3,740	14,510	40,050	55,980	47,475	50,391	36,028	16,026	14,724	5,001	6,784	3,446	3,497	297,652	–
$\sum f_i \cdot (\lg(x_i))^2$	4,663	19,140	55,317	80,351	70,433	76,946	56,436	25,683	24,088	8,337	11,505	5,938	6,115	444,952	–
$\sum f_{i,j} \cdot y_j \cdot \lg(x_i)$	66,26	301,48	919,06	1377,88	1190,73	1291,29	943,23	423,89	388,46	136,94	188,59	91,49	94,60	7413,90	–
$\widetilde{y}_i$	19,9	21,7	23	24	24,8	25,4	25,9	26,4	26,8	27,1	27,4	27,6	27,9		–
$\sum f_{i,j} \cdot (y_j - \widetilde{y}_i)^2$	17	27,5	102,8	148,3	81,5	37,7	32,4	14,4	5,6	0,9	5,4	6,7	1,4	481,6	–

На рис. 31 изображено полученное регрессионное уравнение логарифмической кривой.

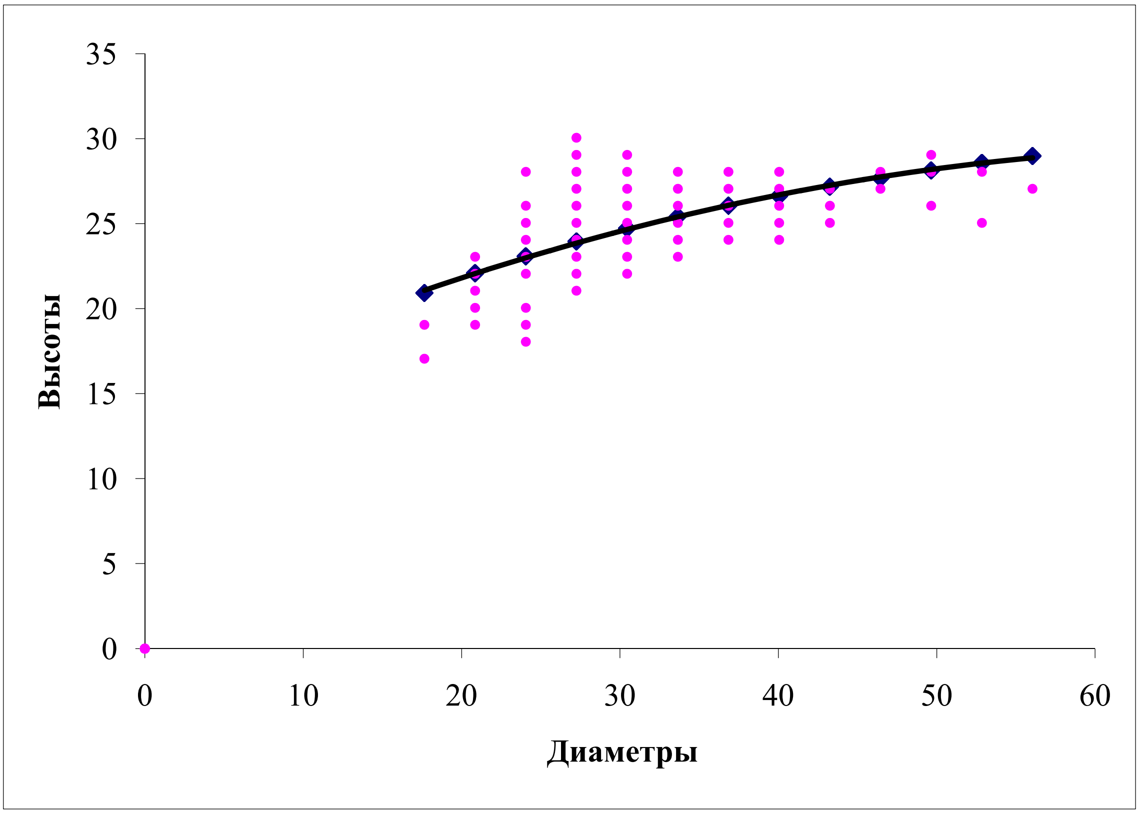


Рис. 31. Зависимость между высотами и диаметрами деревьев в древостое (логарифмическая кривая)

Проанализируем полученное уравнение. Для начала рассчитаем таблицу дисперсионного анализа (табл. 51). Число степеней свободы для суммы квадратов, обусловленной регрессией, в случае логарифмического уравнения равно 1, так как, помимо свободного члена, в нем есть только один коэффициент регрессии – a_1. Общей скорректированной сумме квадратов соответствует число степеней свободы, на единицу меньшее общего количества наблюдений, т. е. $200 - 1 = 199$. На сумму квадратов отклонений относительно регрессии остается $199 - 1 = 198$ степеней свободы. Пользуясь полученным регрессионным уравнением логарифмической кривой (219), определим теоретические высоты $\tilde{y}_i$ и сумму квадратов отклонений эмпирических высот от теоретических (табл. 50). Это значение 526,18 является суммой квадратов отклонений относительно регрессии, или остатком (табл. 51).

Таблица 51. **Дисперсионный анализ уравнения логарифмической кривой**

Источник вариации	Число степеней свободы	Суммы квадратов отклонений *SS*	Средние квадраты *MS*
Обусловленный регрессией	1	525,33	525,33
Относительно регрессии (остаток)	198	526,18	2,66
Общий, скорректированный	199	1051,51	–

Общая скорректированная сумма квадратов отклонений – это сумма квадратов отклонений всех высот от средней арифметической высоты. Данная величина была получена ранее для рассматриваемых данных при вычислении показателей вариации (табл. 23). Сумму квадратов отклонений, обусловленную регрессией, получим как разницу между общей скорректированной суммой квадратов отклонений и остатком: $1051{,}51 - 526{,}18 = 525{,}33$. Разделив суммы квадратов на соответствующие числа степеней свободы, получим средний квадрат, обусловленный регрессией (525,33), и средний квадрат относительно регрессии (2,66).

Теперь рассчитаем F-статистику Фишера, воспользовавшись формулой (141):

$$F = \frac{MS_R}{s^2} = \frac{525{,}33}{2{,}66} = 197{,}49 .$$

С помощью полученного значения проверим гипотезу об отсутствии связи между зависимой и независимой переменными. Для этого в табл. 8 приложения найдем квантиль распределения Фишера с 1 и 198 степенями свободы для уровня значимости $\alpha = 0{,}05$. Это значение равно $F_{0{,}05,1,198} = 3{,}84$. Так как вычисленное значение F-критерия Фишера превышает табличное значение, мы отвергаем нулевую гипотезу об отсутствии связи между высотами и диаметрами. Следовательно, результаты расчетов подтверждают наличие связи между диаметрами и высотами деревьев в древостое, описываемой логарифмической кривой.

Наряду с дисперсионным анализом, вычислим коэффициент детерминации, пользуясь формулой (142)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{525,33}{1051,51} = 0,4996;$$

приведенную R^2 статистику с помощью выражения (143)

$$R_a^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1} =$$

$$= 1 - (1 - 0,4996) \cdot \frac{200-1}{200-1-1} = 0,4971$$

и коэффициент корреляции, который является квадратным корнем из коэффициента детерминации:

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{\frac{525,33}{1051,51}} = 0,7068.$$

Полученные значения последних показателей позволяют оценить степень адекватности логарифмической модели связи между высотами и диаметрами экспериментальным данным как высокую.

Следующий этап оценки качества полученного уравнения регрессии будет заключаться в проверке гипотез о равенстве коэффициентов регрессии нулю. Для этого следует вычислить t-статистики Стьюдента для коэффициентов a_0 и a_1 с помощью формул (152) и (147). Сумму квадратов отклонений значений x_i от их средней арифметической величины из числителя формулы (152) можно вычислить с помощью выражений (207) и (208).

Учитывая то, что в случае логарифмической кривой $x_1 = \lg(x)$, формулу (152) можем представить в виде

$$t_{a_0} = \frac{a_0 \cdot \left(n \cdot \sum_{i=1}^{n}\left(\lg(x_i) - \overline{\lg(x)}\right)^2\right)^{1/2}}{\left(s^2 \cdot \sum_{i=1}^{n}\lg(x_i)^2\right)^{1/2}},$$

а с учетом (207) и (208) сумму квадратов отклонений из числителя полученной формулы преобразуем к выражению

$$\sum_{i=1}^{n}(\lg(x_i)-\overline{\lg(x)})^2 = n\cdot\left(\frac{\sum_{i=1}^{n}\lg(x_i)^2}{n}-\left(\frac{\sum_{i=1}^{n}\lg(x_i)}{n}\right)^2\right)=$$

$$=\sum_{i=1}^{n}\lg(x_i)^2-\frac{\left(\sum_{i=1}^{n}\lg(x_i)\right)^2}{n}. \qquad (220)$$

Теперь, подставляя необходимые суммы из вспомогательной табл. 50 в выражение (220), вычислим значение суммы квадратов отклонений:

$$\sum_{i=1}^{n}(\lg(x_i)-\overline{\lg(x)})^2 =$$

$$=444{,}952-\frac{(297{,}652)^2}{200}=444{,}952-442{,}984=1{,}968. \qquad (221)$$

Далее, подставив полученное значение, остаточный средний квадрат s^2 из табл. 51 и сумму из табл. 50 в формулу (221), вычислим критерий Стьюдента для коэффициента регрессии a_0:

$$t_{a_0}=\frac{a_0\cdot\left(n\cdot\sum_{i=1}^{n}(\lg(x_i)-\overline{\lg(x)})^2\right)^{1/2}}{\left(s^2\cdot\sum_{i=1}^{n}\lg(x_i)^2\right)^{1/2}}=$$

$$=\frac{0{,}8999\cdot(200\cdot 1{,}968)^{1/2}}{(2{,}66\cdot 444{,}952)^{1/2}}=\frac{17{,}8534}{34{,}4031}=0{,}5189.$$

Аналогичным образом, с учетом того, что в случае логарифмической кривой $x_1 = \lg(x)$, преобразуем формулу (147), как показано ниже:

$$t_{a_1}=a_1\cdot\left(\frac{\sum_{i=1}^{n}(x_{1,i}-\bar{x}_1)^2}{s^2}\right)^{1/2}=a_1\cdot\left(\frac{\sum_{i=1}^{n}(\lg(x_i)-\overline{\lg(x)})^2}{s^2}\right)^{1/2}. \qquad (222)$$

Теперь, подставив в выражение (222) вычисленную ранее сумму квадратов отклонений (221) и остальные значения, определим t-статистику Стьюдента для коэффициента a_1:

$$t_{a_1} = a_1 \cdot \left(\frac{\sum_{i=1}^{n} (\lg(x_i) - \overline{\lg(x)})^2}{s^2} \right)^{1/2} = 16{,}06 \cdot \left(\frac{1{,}968}{2{,}66} \right)^{1/2} = 13{,}81 .$$

Далее, для проверки гипотезы о равенстве коэффициентов регрессии a_0 и a_1 нулю, в табл. 5 приложения следует найти значение квантиля t-распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и 198 степеней свободы $t_{0,05;198} = 1{,}653$. Так как абсолютное значение вычисленной t-статистики для коэффициента a_1 больше табличного $\left| t_{a_1} \right| = |13{,}81| > 1{,}653$, то нулевую гипотезу о равенстве нулю этого коэффициента регрессии следует отвергнуть. Следовательно, коэффициент регрессии a_1 значим. Что касается a_0, то в отношении этого коэффициента у нас нет оснований отвергать нулевую гипотезу о равенстве данного регрессионного коэффициента нулю, так как абсолютное значение вычисленной t-статистики для коэффициента a_1 меньше табличного $\left| t_{a_0} \right| = |0{,}5189| < 1{,}653$. Мы принимаем нулевую гипотезу. Следовательно, коэффициент регрессии a_0 не значим.

Стандартная ошибка регрессионного уравнения логарифмической кривой может быть вычислена по формуле (140), но в связи с тем, что исходные данные в рассматриваемом примере сгруппированы и вычисления ведутся на основе корреляционной решетки (табл. 14), удобнее воспользоваться следующим вариантом формулы:

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} .$$

Подставляя в формулу (189) необходимые значения из табл. 50 и учитывая, что в рассматриваемом примере $m = 1$, получим

$$s = \sqrt{\frac{\sum f_{i,j} \cdot (y_j - \tilde{y}_i)^2}{n - m - 1}} = \sqrt{\frac{526{,}18}{200 - 1 - 1}} = 1{,}630 .$$

Используя данные дисперсионного анализа уравнения регрессии, приведенные в табл. 51, стандартную ошибку оценки можно определить, вычислив квадратный корень из среднего квадрата относительно регрессии:

$$s = \sqrt{2{,}66} = 1{,}631 .$$

7.7. Выбор регрессионной модели

В предыдущих подразделах были рассмотрены примеры регрессионного анализ зависимости между высотами и диаметрами деревьев в чистом одновозрастном древостое. В этих примерах использовались разные модели связи: линейная, параболическая, гиперболическая и логарифмическая. Характерно, что результаты расчетов статистически достоверно подтверждают наличие связи между диаметрами и высотами деревьев в древостое для всех рассмотренных видов моделей. В таких ситуациях встает вопрос, какую модель выбрать? Для анализа соответствия регрессионных моделей экспериментальным данным в рассматриваемых примерах был рассчитан ряд статистик, характеризующих полученные уравнения. Сведем эти параметры в табл. 52 и проанализируем их.

Нетрудно заметить, что наилучшим образом описывает связь между диаметрами и высотами в древостое уравнение гиперболы. Об этом говорит максимальное значение коэффициента корреляции 0,7415. Величина коэффициента детерминации показывает, что данное уравнение описывает 54,99% всей вариации высот, что больше, чем у других моделей. Стандартная ошибка данного уравнения также меньше, чем у остальных уравнений, для которых выполнялись расчеты. В пользу уравнения гиперболы говорят также и самые большие величины приведенной R^2 статистики и критерия Фишера.

Таблица 52. **Статистические показатели, характеризующие уравнения регрессии**

Уравнение	Коэффициент корреляции R	Коэффициент детерминации R^2	Преобразованный коэффициент детерминации R_a^2	Критерий Фишера F	Стандартная ошибка оценки S^2
$h = a_0 + a_1 \cdot d$	0,6578	0,4327	0,4298	151,16	1,735
$h = a_0 + a_1 \cdot d + a_2 \cdot d^2$	0,7383	0,5451	0,5405	117,95	1,559
$h = a_0 + a_1 / d$	0,7415	0,5499	0,5476	241,93	1,546
$h = a_0 + a_1 \ln(d)$	0,7068	0,4996	0,4971	197,49	1,630

Таким образом, лучшей моделью, характеризующей связь высот и диаметров деревьев в древостое, является уравнение гиперболы:

$$h = 32{,}35 - \frac{226{,}3}{d};\ t_{a_0} = 65{,}29;\ t_{a_1} = -15{,}61.$$

В рассмотренном варианте никаких проблем с выбором лучшего уравнения нет. Все характеристики указывают на гиперболу. Но так бывает не всегда. В некоторых случаях выбрать лучшую регрессионную модель из рассмотренных вариантов достаточно сложно.

Предположим, что анализировались три модели связи: линейная, параболическая и логарифмическая (исключим гиперболу из рассмотрения) (табл. 53).

Таблица 53. **Статистические показатели, характеризующие уравнения регрессии**

Уравнение	Коэффициент корреляции R	Коэффициент детерминации R^2	Преобразованный коэффициент детерминации R_a^2	Критерий Фишера F	Стандартная ошибка оценки S^2
$h = a_0 + a_1 \cdot d$	0,6578	0,4327	0,4298	151,16	1,735
$h = a_0 + a_1 \cdot d + a_2 \cdot d^2$	0,7383	0,5451	0,5405	117,95	1,559
$h = a_0 + a_1 \ln(d)$	0,7068	0,4996	0,4971	197,49	1,630

В таком варианте коэффициент корреляции, коэффициент детерминации, стандартная ошибка оценки указывают на параболу второго порядка. Именно она описывает больше всего вариации высот деревьев и позволяет с наименьшей ошибкой оценивать величину высоты дерева на основании измеренного диаметра. В том случае, если такая оценка и является целью разработки модели, данный выбор является правильным. Вместе с тем у параболы второго порядка самый маленький критерий Фишера. Дело в том, что этот показатель учитывает число степеней свободы. В рассматриваемом примере парабола является самой сложной моделью в связи с тем, что она имеет три коэффициента регрессии – a_0, a_1 и a_2, а не два, как другие уравнения. Следовательно, на уравнение параболы второго порядка приходится на одну степень свободы больше, чем на другие уравнения, что и отражается на величине критерия Фишера. Уменьшение стандартной ошибки оценки и увеличение доли вариации высот, описываемой уравнением параболы второго порядка, в сравнении с уравнениями прямой линии или логарифмической кривой, является незначительным на фоне увеличения в полтора раза количества параметров уравнения. Именно поэтому критерий Фишера, вычисленный для параболы, оказался самым маленьким. В такой ситуации иногда целесообразно отдать предпочтение более простому уравнению, например логарифмическому, несмотря на то, что у него стандартная ошибка оценки несколько выше (1,630), чем у параболы второго порядка (1,559). Так можно поступить в том случае, если необходимо построить достаточно простую модель связи и когда оценка высот деревьев на основании их диаметров не является основной целью моделирования.

Приведенная R^2 статистика также учитывает число степеней свободы, однако в значительно меньшей степени, чем критерий Фишера. В рассматриваемом примере, в отличие от F-критерия, она характеризует анализируемые уравнения так же, как и коэффициент детерминации.

ПРИЛОЖЕНИЕ

Таблица 1. **Соотношения диаметров и высот деревьев в спелых сосновых насаждениях**

Номера столбцов															
1				2				3				4			
D	*H*	*D*	*D*	*H*	*D*	*H*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*
28,2	24,6	35,0	35,4	25,4	27,0	24,6	25,7	37,0	25,5	28,2	24,0	35,4	25,4	27,0	24,6
30,6	22,2	41,7	28,6	24,3	32,0	25,6	26,0	28,5	25,5	25,5	23,5	28,6	24,3	32,0	25,6
20,6	21,7	52,3	45,1	28,2	22,0	23,9	24,5	20,1	20,2	25,5	23,5	45,1	28,2	22,0	23,9
26,7	23,7	29,6	30,1	25,6	53,4	27,8	26,1	40,5	28,3	29,3	23,5	30,1	25,6	53,4	27,8
26,5	23,5	20,5	28,3	24,4	29,3	24,9	26,6	29,2	26,5	33,4	26,0	28,3	24,4	29,3	24,9
27,6	25,6	52,1	34,6	28,4	26,5	24,3	27,5	42,4	25,5	38,5	25,1	34,6	28,4	26,5	24,3
30,5	24,5	34,6	41,0	26,4	27,5	25,9	25,4	34,6	25,6	38,5	25,3	41,0	26,4	27,5	25,9
26,1	22,6	28,5	31,0	24,6	37,0	23,9	27,6	30,5	23,8	22,5	21,5	31,0	24,6	37,0	23,9
29,9	22,2	35,0	25,6	22,2	37,5	25,5	26,6	21,5	22,9	28,5	25,2	25,6	22,2	37,5	25,5
21,0	21,6	35,1	27,5	24,0	35,9	23,9	26,1	48,5	28,7	25,6	25,5	27,5	24,0	35,9	23,9
44,0	28,6	27,5	30,1	24,5	37,6	28,8	24,0	24,7	23,7	35,7	25,6	30,1	24,5	37,6	28,8
38,5	26,5	40,6	25,6	23,8	26,6	24,4	24,1	33,8	26,4	27,5	25,5	25,6	23,8	26,6	24,4
31,5	25,6	35,6	30,5	23,1	25,5	23,3	25,6	25,5	24,5	29,5	23,5	30,5	23,1	25,5	23,3
28,6	21,1	24,6	25,1	25,4	26,5	23,5	28,6	22,5	21,5	21,8	21,5	25,1	25,4	26,5	23,5
33,0	26,6	27,0	31,5	24,5	27,6	24,6	22,7	25,0	25,5	27,9	24,2	31,5	24,5	27,6	24,6
28,1	22,3	41,2	22,6	21,7	37,0	25,9	28,4	33,2	27,9	34,1	25,5	22,6	21,7	37,0	25,9
38,6	25,1	47,5	26,5	23,4	38,0	27,5	20,2	23,3	21,5	36,2	24,2	26,5	23,4	38,0	27,5
33,5	26,0	27,6	34,5	25,6	31,9	24,9	24,6	34,5	27,5	37,3	26,3	34,5	25,6	31,9	24,9
31,0	24,5	39,0	36,0	28,4	28,0	25,9	26,0	42,5	25,5	32,5	24,5	36,0	28,4	28,0	25,9
23,0	23,6	39,9	18,6	20,4	34,6	26,0	27,0	40,5	28,3	37,5	25,5	18,6	20,4	34,6	26,0
25,0	21,6	39,0	25,8	22,3	23,0	22,1	27,6	37,6	24,5	24,5	24,1	25,8	22,3	23,0	22,1
32,0	25,0	19,0	26,9	21,3	27,0	25,0	25,6	38,5	26,7	33,6	24,0	26,9	21,3	27,0	25,0
37,6	27,0	27,5	30,9	26,3	26,0	22,5	23,6	24,7	22,2	41,5	25,5	30,9	26,3	26,0	22,5
43,0	27,0	37,0	26,2	24,6	29,2	22,7	29,6	23,5	21,5	31,5	26,5	26,2	24,6	29,2	22,7
29,6	21,6	17,5	39,0	26,9	26,6	24,5	25,4	17,4	18,0	28,5	22,0	39,0	26,9	26,6	24,5

Продолжение табл. 1

Номера столбцов															
5				6				7				8			
D	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*
28,5	25,0	23,0	23,5	23,6	20,1	22,6	21,5	43,0	28,4	21,5	27,1	27,0	22,0	38,0	25,7
20,0	22,5	38,5	25,0	31,0	23,6	22,6	22,6	46,4	26,1	33,6	24,0	49,6	27,9	30,5	23,6
22,5	22,5	22,0	21,5	32,4	25,7	41,0	27,3	60,0	27,6	22,3	22,6	22,1	23,2	27,0	23,0
28,0	24,5	37,2	25,0	36,6	26,4	31,0	24,5	37,0	26,4	27,0	23,6	35,6	26,1	33,1	26,7
30,5	22,0	33,3	25,5	25,6	23,9	28,0	25,9	24,0	24,4	27,3	25,4	39,1	26,9	25,1	23,6
19,5	23,0	46,4	26,5	29,6	24,7	26,6	23,0	37,6	27,7	31,6	27,4	31,5	25,0	33,1	24,5
30,0	25,0	25,5	22,0	34,4	27,4	28,0	25,1	38,3	25,5	29,5	24,0	33,3	25,5	23,6	21,4
22,0	23,0	28,0	23,5	29,0	23,9	31,0	25,4	31,5	24,6	30,0	25,5	32,0	24,6	21,4	22,0
52,0	26,5	26,5	24,0	27,4	22,0	38,0	24,5	31,0	25,5	22,6	22,9	27,3	23,9	28,6	25,4
48,0	28,0	36,0	26,0	44,9	28,4	23,6	20,0	32,5	25,0	20,0	22,7	40,1	27,1	30,2	25,3
37,0	26,5	25,5	23,5	28,3	24,1	31,0	23,5	20,6	20,0	37,3	25,5	34,5	27,8	27,0	25,4
21,5	21,0	31,0	25,5	45,0	26,5	32,4	25,6	25,6	24,5	44,5	28,4	38,5	25,6	36,4	26,6
26,5	24,5	26,5	21,5	36,5	24,1	36,6	26,6	26,6	25,5	41,2	26,4	26,4	25,6	23,6	23,0
21,0	21,5	30,5	24,5	30,4	26,1	25,5	23,6	21,5	22,0	31,5	25,9	22,6	22,8	38,6	28,0
36,0	26,5	46,5	24,6	38,8	27,6	29,0	24,5	20,5	22,5	60,0	28,5	51,6	27,5	24,6	24,9
24,5	25,0	34,5	25,5	41,4	26,7	34,1	27,1	28,5	24,0	25,3	21,7	26,6	23,0	29,6	23,7
37,0	27,5	23,5	21,5	39,1	27,6	29,4	23,6	21,5	22,2	41,0	27,5	26,6	22,5	38,7	26,4
26,0	25,5	29,5	24,5	31,6	24,1	27,1	22,2	29,5	24,0	27,0	23,6	41,0	27,6	42,6	25,9
22,5	22,5	27,0	22,0	22,6	22,9	44,5	28,1	22,6	24,4	29,0	24,0	25,0	23,0	32,7	25,0
35,0	25,0	25,5	25,0	23,0	23,7	28,0	24,1	24,0	20,3	29,5	25,6	33,1	27,5	48,8	29,0
28,0	25,5	22,5	24,0	33,4	26,5	45,0	26,5	30,7	23,9	29,6	27,0	22,5	24,0	23,0	24,5
36,5	25,0	29,5	27,0	22,4	22,6	36,6	24,0	25,0	24,0	37,1	27,5	37,1	25,6	26,6	22,7
51,0	27,5	32,0	23,5	22,5	22,1	41,0	27,6	42,5	27,6	45,0	27,0	40,5	27,5	35,1	24,7
21,0	22,0	51,0	27,5	27,3	24,4	53,0	27,6	28,3	25,3	28,3	24,1	27,0	23,0	29,6	24,9
36,5	25,5	24,0	23,5	17,0	18,3	31,0	24,0	17,6	18,5	33,2	26,6	34,4	26,7	31,8	24,4

Продолжение табл. 1

Номера столбцов															
9				10				11				12			
D	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*
36,0	26,6	25,0	23,6	20,6	21,4	41,6	27,6	20,1	22,7	31,0	23,5	23,0	23,6	26,0	23,6
44,0	30,3	34,5	24,6	31,3	23,6	40,2	27,2	29,6	21,8	29,6	22,9	26,1	22,6	35,2	24,4
36,4	24,4	28,0	25,6	27,6	25,6	32,0	24,6	18,3	19,9	27,5	23,8	38,0	25,6	23,4	23,5
24,3	23,3	24,2	21,5	31,5	25,9	25,0	24,6	30,5	24,2	32,0	26,5	34,8	24,9	29,3	24,1
32,2	24,7	20,3	22,7	29,6	25,0	23,0	21,5	36,5	26,7	42,0	25,6	22,6	20,7	32,2	25,1
39,5	26,1	27,5	22,6	30,0	25,9	37,5	26,6	27,3	24,8	31,5	23,5	25,6	24,4	34,6	27,1
50,5	28,1	36,0	28,1	28,5	25,4	46,6	26,6	38,5	27,7	27,1	22,0	52,1	29,6	40,1	25,8
40,6	25,4	43,0	28,0	27,0	24,4	39,6	26,1	37,6	26,4	26,7	24,3	49,6	28,5	28,6	23,9
25,6	25,0	37,1	25,6	35,5	25,4	46,0	26,6	31,6	25,8	29,5	23,6	26,3	23,7	25,9	23,8
27,0	22,0	36,6	26,1	38,0	25,1	31,5	26,6	45,1	28,8	31,0	25,4	37,0	24,1	32,6	25,8
22,0	21,5	32,0	24,6	37,3	25,6	21,6	21,6	46,9	29,0	32,0	24,5	25,8	22,7	27,0	23,7
29,5	27,6	37,0	25,2	16,0	19,5	31,0	24,6	46,6	28,7	22,6	23,1	17,0	18,7	22,6	22,8
28,4	27,0	27,6	29,6	32,0	24,6	39,0	23,5	20,0	21,1	32,6	25,8	26,5	23,9	31,7	23,8
35,0	24,5	33,5	25,6	34,6	25,6	32,0	24,5	27,0	22,4	45,5	28,3	47,3	28,8	24,9	23,6
42,0	25,0	31,0	25,1	35,1	25,4	21,5	23,3	36,4	24,7	23,6	22,2	37,6	24,7	43,8	28,1
36,6	27,1	24,6	23,6	24,6	23,9	31,0	26,0	20,5	22,8	25,0	23,7	20,9	21,5	38,6	27,1
34,1	25,7	24,0	21,6	18,6	21,6	33,0	26,6	32,3	25,6	33,0	26,4	36,1	24,9	27,0	22,5
27,1	23,1	23,1	22,1	27,6	22,7	44,4	26,6	29,5	23,6	35,4	24,9	19,5	19,8	21,1	23,6
23,6	22,5	25,1	24,5	46,6	28,3	24,6	22,7	26,0	23,7	36,3	25,4	27,0	22,9	32,0	25,7
44,1	28,0	53,4	27,6	24,4	22,2	28,5	24,0	28,2	26,0	30,0	24,5	36,2	25,8	40,3	27,6
39,7	23,9	45,0	26,5	32,2	24,4	32,0	23,5	22,5	22,7	31,5	24,6	23,1	21,8	32,0	25,8
21,5	20,0	21,5	21,0	27,0	23,9	29,6	27,4	34,3	27,4	39,6	26,6	29,0	23,8	29,0	23,6
29,0	23,6	36,6	25,6	26,0	21,0	39,6	26,0	30,6	23,7	23,3	22,6	41,2	28,2	31,7	23,6
32,1	24,6	28,5	25,4	23,6	25,9	28,6	25,5	24,1	23,5	41,0	26,0	31,6	24,4	30,9	23,5
29,6	24,6	31,1	24,4	24,0	20,1	43,0	26,6	42,6	26,7	34,5	26,6	30,2	24,1	30,7	23,4

Продолжение табл. 1

Номера столбцов															
13				14				15				16			
D	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*
46,6	26,4	29,6	27,1	47,0	27,7	28,5	25,2	43,4	28,4	52,4	27,1	28,1	24,4	32,0	25,4
21,6	20,1	29,7	24,6	35,3	22,5	30,3	25,2	20,1	20,4	44,6	28,6	30,1	24,3	39,6	26,9
34,0	27,0	26,6	25,5	36,3	24,6	33,3	25,2	26,9	25,7	31,6	24,7	30,4	24,1	28,1	26,4
30,6	23,1	25,5	23,7	22,0	22,4	36,5	26,6	29,3	23,6	23,6	23,0	34,6	26,3	29,6	25,1
44,0	27,3	23,1	21,6	48,5	26,1	37,6	26,6	21,0	21,5	26,5	24,4	22,6	17,7	49,6	25,7
38,1	24,5	29,5	23,5	35,3	26,2	26,5	23,3	30,4	22,4	22,6	22,7	42,2	25,0	34,6	25,7
40,5	27,1	22,6	22,2	38,0	25,5	17,5	18,6	28,4	25,3	31,0	24,6	26,6	23,6	30,1	25,6
26,1	22,6	42,0	26,6	22,2	24,0	46,4	24,6	17,6	18,6	29,5	23,5	41,0	27,0	24,6	22,7
24,2	22,1	30,6	25,6	23,3	29,0	29,0	26,5	30,3	25,1	28,9	25,4	37,1	27,6	27,5	25,7
31,6	27,3	41,6	25,6	38,5	26,2	38,5	25,5	26,0	23,7	32,6	24,7	37,0	26,4	35,1	24,6
46,6	24,6	23,0	19,0	28,3	24,5	26,2	25,6	35,4	25,5	31,0	25,4	25,1	24,4	20,1	23,1
24,5	24,6	44,0	28,3	24,0	23,6	15,6	19,1	33,0	25,7	24,6	22,6	22,0	20,1	24,7	22,2
39,0	27,6	21,5	22,2	20,0	22,6	37,5	28,6	24,6	22,7	50,4	28,5	24,6	25,9	27,1	24,7
28,0	22,2	36,5	27,6	25,5	23,5	40,1	27,3	39,9	26,4	32,0	24,1	30,6	23,0	22,1	22,7
36,0	27,0	26,0	21,6	22,0	22,2	19,0	18,0	31,0	23,5	22,6	17,9	21,0	22,6	49,5	27,5
33,3	25,0	28,0	23,5	30,5	23,3	15,5	19,0	31,4	25,6	27,0	22,0	33,0	24,6	22,5	22,0
24,6	23,6	28,4	22,7	29,6	22,5	26,0	22,5	39,6	27,0	19,3	22,4	18,6	20,0	26,5	23,0
39,6	26,6	36,5	26,0	47,4	27,5	30,6	23,6	28,0	22,6	29,4	24,9	24,0	22,0	34,6	27,6
33,6	26,3	40,5	25,5	17,5	19,6	29,6	25,2	24,6	22,7	32,0	23,5	44,4	26,6	28,0	22,7
29,5	24,6	40,6	26,5	33,0	26,0	33,3	26,0	38,7	25,6	40,6	27,5	25,1	23,6	23,4	22,9
32,0	24,3	17,5	19,6	26,0	25,5	28,0	25,0	25,9	23,6	29,6	23,4	42,4	26,6	37,4	23,8
39,3	26,3	29,6	21,5	27,0	23,6	32,5	19,4	22,0	23,9	28,0	23,5	36,6	26,7	25,1	24,1
41,0	27,6	36,5	26,0	25,0	23,6	40,0	26,6	41,6	27,6	30,5	26,0	28,7	24,4	26,1	22,6
37,6	27,6	35,1	24,6	28,0	22,6	30,3	25,7	30,6	24,0	40,6	24,1	18,7	17,9	30,9	25,6
36,0	24,4	23,1	22,0	37,3	26,2	22,9	17,8	28,4	24,4	22,9	22,6	31,0	23,7	21,6	21,9

Окончание табл. 1

Номера столбцов															
17				18				19				20			
D	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*	*H*
25,5	24,6	24,0	21,6	28,1	25,6	29,5	24,5	46,6	27,7	33,5	25,6	36,4	26,6	27,0	23,0
42,1	25,9	28,3	25,3	36,0	28,0	22,2	21,9	28,0	25,3	36,5	25,0	27,5	23,4	34,5	24,6
28,1	25,7	36,6	26,6	26,4	22,6	33,6	27,3	32,6	25,7	31,6	25,5	30,6	23,6	25,0	23,7
32,6	24,6	24,5	22,9	56,6	26,6	31,5	27,0	28,6	26,6	41,0	27,0	27,0	23,0	32,0	24,0
22,7	22,4	33,0	27,4	38,6	28,0	39,5	26,6	22,6	24,0	31,0	24,8	20,4	20,3	22,6	22,7
27,7	25,6	36,0	28,4	36,6	26,0	25,0	23,6	35,6	25,7	37,0	24,0	38,0	27,5	33,5	25,6
27,0	25,1	56,6	26,7	30,1	24,6	40,6	25,3	33,5	25,5	26,6	29,0	24,6	27,6	26,4	29,6
45,1	26,6	25,0	23,5	35,2	22,6	36,0	25,6	43,5	26,0	27,0	23,6	19,9	19,4	36,6	24,4
37,5	26,0	33,7	23,4	33,3	26,6	22,6	18,6	30,3	24,4	21,9	20,1	32,6	25,9	24,6	22,7
26,6	22,6	22,6	17,7	30,3	28,3	43,4	27,5	24,0	22,6	38,5	26,5	37,5	26,7	22,0	22,7
32,7	25,9	35,0	25,6	19,4	19,3	26,5	24,0	35,6	25,6	36,5	26,5	34,6	25,6	38,4	25,5
42,7	27,3	33,4	27,4	42,6	26,6	30,6	26,0	28,1	24,3	31,0	25,0	34,6	26,4	26,5	24,0
27,0	24,3	27,1	23,9	28,1	24,7	25,6	22,4	24,4	24,4	30,5	26,0	30,6	23,9	35,0	24,9
24,6	22,3	26,7	28,4	31,7	25,8	28,7	21,3	30,5	23,6	51,0	27,7	20,6	20,4	51,8	24,8
25,7	23,7	31,0	22,0	37,8	26,0	32,6	24,8	29,6	22,7	30,7	23,0	22,6	24,0	33,0	26,4
42,0	25,6	31,6	26,6	26,3	22,7	35,1	24,4	28,5	21,6	24,5	25,7	39,0	27,9	30,0	24,4
35,1	24,7	38,6	26,0	26,6	23,0	31,0	24,5	32,6	26,4	38,0	25,6	20,6	20,4	22,5	21,7
24,5	22,6	28,0	23,5	27,3	23,0	26,0	22,6	26,0	24,7	36,6	25,7	33,4	25,0	26,7	25,9
28,6	25,9	53,0	27,6	22,5	22,7	34,6	25,8	22,2	21,6	23,0	22,7	26,0	22,9	30,5	28,9
49,5	25,7	25,8	24,7	38,0	27,5	31,4	24,8	25,7	23,3	22,7	22,5	32,4	24,9	32,7	26,0
30,6	25,0	39,5	26,0	20,5	21,6	28,0	25,6	17,7	18,8	29,5	24,4	22,7	24,0	18,6	17,5
50,0	28,6	29,5	24,0	42,2	25,0	38,1	27,7	32,1	25,3	34,5	25,5	34,8	25,4	30,6	23,6
37,0	23,7	32,0	24,6	39,6	24,4	27,6	25,5	37,2	26,2	23,8	24,4	42,0	25,7	42,6	26,6
21,7	20,7	32,6	24,7	26,6	25,6	29,9	27,0	38,5	26,7	49,5	27,7	30,6	23,8	22,5	22,0
45,0	26,7	31,6	25,7	30,2	28,3	33,3	28,4	23,6	20,2	32,6	25,9	18,6	17,5	24,0	22,0

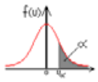
Таблица 2. **Нормированное нормальное распределение**

Функция распределения: $\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$

X	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,504	0,506	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,728	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,866	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,950	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,961	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,996	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999
3,0	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,1	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,2	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	1,000
3,3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
3,4	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

1.2

Таблица 3. **Нормированное нормальное распределение**

Значения U_α в зависимости от вероятности $\alpha : P\{U > U_\alpha\} = \alpha$ 

α	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0 005
U_α	0,84	1,28	1,64	1,96	2,33	2,58	3,09	3,29

Таблица 4. **Квантили распределения Колмогорова λ_α**

В таблице приведены значения квантилей λ_α в зависимости от вероятности α такой, что $P(\lambda \geq \lambda_\alpha) = \alpha$.

α	λ_α	α	λ_α	α	λ_α
0,99	0,44	0,50	0,83	0,15	1,14
0,90	0,57	0,40	0,89	0,10	1,22
0,80	0,64	0,30	0,97	0,05	1,36
0,70	0,71	0,25	1,02	0,02	1,53
0,60	0,77	0,20	1,07	0,01	1,63

Таблица 5. **Распределение Стьюдента**

Значения $t_{\alpha,\gamma}$ удовлетворяют условию

$$P\left(t \geq t_{\alpha,\gamma}\right) = \int_{t_{\alpha,\gamma}}^{\infty} S(t,\gamma)dt = \alpha .$$

В таблице приведены значения квантилей $t_{\alpha,\gamma}$ в зависимости от числа степеней свободы γ и вероятности α.

γ \ α	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,614	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	5,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,288	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	3,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	0,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	5,042	2,457	2,750	3,385	3,646
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Таблица 6. **Доверительные интервалы для σ**

Нижняя ν_1 и верхняя ν_2 границы доверительного интервала

$$\nu_1 \cdot \tilde{\sigma} < \sigma < \nu_2 \cdot \tilde{\sigma},\ \tilde{\sigma} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

P / $\nu = n-1$	0,99		0,98		0,95		0,90	
	ν_1	ν_2	ν_1	ν_2	ν_1	ν_2	ν_1	ν_2
1	0,356	159	0,388	79,8	0,446	31,9	0,510	15,9
2	0,434	14,1	0,466	9,97	0,521	6,28	0,578	4,40
3	0,483	6,47	0,514	5,11	0,566	3,73	0,620	2,92
4	0,519	4,39	0,549	3,67	0,599	2,87	0,649	2,37
5	0,546	3,48	0,576	3,00	0,624	2,45	0,672	2,090
6	0,569	2,98	0,597	2,62	0,644	2,202	0,690	1,916
7	0,588	2,66	0,616	2,377	0,661	2,035	0,705	1,797
8	0,604	2,440	0,631	2,205	0,675	1,916	0,718	1,711
9	0,618	2,277	0,644	2,076	0,688	1,826	0,729	1,645
10	0,630	2,154	0,656	1,977	0,699	1,755	0,739	1,593
11	0,641	2,056	0,667	1,898	0,708	1,698	0,748	1,550
12	0,651	1,976	0,677	1,833	0,717	1,651	0,755	1,515
13	0,660	1,910	0,685	1,779	0,725	1,611	0,762	1,485
14	0,669	1,854	0,693	1,733	0,732	1,577	0,769	1,460
15	0,676	1,806	0,700	1,694	0,739	1,548	0,775	1,437
16	0,683	1,764	0,707	1,659	0,745	1,522	0,780	1,418
17	0,690	1,727	0,713	1,629	0,750	1,499	0,785	1,400
18	0,696	1,695	0,719	1,602	0,756	1,479	0,790	1,385
19	0,702	1,666	0,725	1,578	0,760	1,460	0,794	1,370
20	0,707	1,640	0,730	1,556	0,765	1,444	0,798	1,358
21	0,712	1,617	0,734	1,536	0,769	1,429	0,802	1,346
22	0,717	1,598	0,739	1,519	0,773	1,416	0,805	1,335
23	0,722	1,576	0,743	1,502	0,777	1,402	0,809	1,326
24	0,726	1,558	0,747	1,487	0,781	1,391	0,812	1,316
25	0,730	1,541	0,751	1,473	0,784	1,380	0,815	1,308
26	0,734	1,526	0,755	1,460	0,788	1,371	0,818	1,300
27	0,737	1,512	0,758	1,448	0,791	1,361	0,820	1,293
28	0,741	1,499	0,762	1,436	0,794	1,352	0,823	1,286
29	0,744	1,487	0,765	1,426	0,796	1,344	0,825	1,279
30	0,748	1,475	0,768	1,417	0,799	1,337	0,828	1,274
40	0,774	1,390	0,792	1,344	0,821	1,279	0,847	1,228
50	0,793	1,336	0,810	1,297	0,837	1,243	0,861	1,199
60	0,808	1,299	0,824	1,265	0,489	1,217	0,871	1,179
70	0,820	1,272	0,835	1,241	0,858	1,198	0,879	1,163
80	0,829	1,250	0,844	1,222	0,866	1,183	0,886	1,151
90	0,838	1,233	0,852	1,207	0,873	1,171	0,892	1,141
100	0,845	1,219	0,858	1,195	0,878	1,161	0,897	1,133
200	0,887	1,15	0,897	1,13	0,912	1,11	0,925	1,09

Таблица 7. **Критические значения χ^2 распределения**

В таблице приведены значения квантилей $\chi^2_{\alpha,\gamma}$ в зависимости от числа степеней свободы γ и вероятности α такими, что $P\left(\chi^2 \geq \chi^2_{\alpha,\gamma} = \alpha\right)$.

γ	α											
	0,99	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,01	0,001
1	0,000157	0,00393	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	6,635	10,827
2	0,0201	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	9,210	13,8915
3	0,115	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	11,345	16,266
4	0,297	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	13,277	18,467
5	0,554	1,145	4,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	15,086	20,515
6	0,872	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	16,812	22,457
7	1,239	2,167	2,833	3,822	4,671	6,346	8,363	9,803	12,017	14,067	18,475	24,322
8	1,646	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	20,090	26,125
9	2,088	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	213,666	27,877
10	2,558	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	23,209	29,588
11	3,053	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	24,725	31,264
12	3,571	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,519	21,026	26,207	32,909
13	4,107	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	27,688	34,528
14	4,660	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	29,141	36,123
15	5,229	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	30,578	37,697
16	5,812	7,962	9,312	11,152	012,624	15,338	18,418	20,465	23,542	26,296	32,000	39,252
17	6,408	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	33,409	40,790
18	7,015	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	34,805	42,312
19	7,633	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	36,191	43,820
20	8,260	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	37,566	45,315
21	8,897	11,591	13,240	15,445	17,182	20,337	23,853	26,171	29,615	32,671	38,932	46,797
22	9,542	12,338	14,041	16,310	18,101	21,337	24,939	27,301	30,813	33,924	40,289	48,268
23	10,196	13,091	014,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	41,638	49,728
24	10,856	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	42,980	51,179
25	11,524	14,611	16,473	18,940	20,867	24,337	28,172	30,375	34,382	37,652	44,314	52,620
26	12,198	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	45,642	54,052
27	12,879	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	46,963	55,476
28	13,565	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	48,278	56,893
29	14,256	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	49,588	58,302
30	14,953	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	50,892	59,703

Таблица 8. ***F*- распределение Фишера**

В таблице приведены верхние 1% – точки ***F*** (v_1, v_2, 0,99)

γ_2	Степени свободы для числителя: γ_1																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,48	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10

Продолжение табл. 8

γ_2	Степени свободы для числителя: γ_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
50	7,17	5,06	4,2	3,72	3,41	3,19	3,02	2,89	2,78	2,7	2,56	2,42	2,27	2,18	2,1	2,01	1,91	1,8	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	7,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

В таблице приведены верхние 10% – точки $\boldsymbol{F}$ (v_1, v_2, 0,90)

$v2$	$v1$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06	63,33
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,18	5,18	5,17	5,16	5,15	5,14	5,13
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66

Продолжение табл. 8

19	2,99	2,61	2,40	2,27	2,17	2,11	2,06	2,02	1,98	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63
20	2,97	2,59	2,38	2,25	2,16	2,29	2,04	2,00	1,96	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53
25	2,92	2,53	2,32	2,18	2,09	2,02	197	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38
50	2,81	2,41	2,2	2,06	1,97	1,9	1,84	1,8	1,76	1,73	1,68	1,63	1,57	1,54	1,5	1,46	1,42	1,38	1,33
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00

В таблице приведены верхние 5% – точки $\boldsymbol{F}$ (ν_1, ν_2, 0,95)

ν2	ν1																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,61	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54

Окончание табл. 8

11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,47	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	184	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,82	1,79	1,74	1,69	1,64	1,58	1,51
50	4,03	3,18	2,79	2,56	2,4	2,29	2,2	2,13	2,07	2,03	1,95	1,87	1,78	1,74	1,69	1,63	1,58	1,51	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,29	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Таблица 9. **Значение нормальной функции распределения $f(x)$ и ее производных $f^{III}(x)$ и $f^{IV}(x)$**

x	$f(x)$	$f^{III}(x)$	$f^{IV}(x)$	x	$f(x)$	$f^{III}(x)$	$f^{IV}(x)$
1	2	3	4	1	2	3	4
0,00	0,39 894	+0,00 000	+1,19 683	0,45	0,36 053	+0,45 386	+0,65 832
0,01	0,39 892	+0,01 197	+1,19 653	0,46	0,35 889	+0,46 034	+0,63 709
0,02	0,39 886	+0,02 393	+1,19 563	0,47	0,35 723	+0,46 660	+0,41 564
0,03	0,39 876	+0,03 588	+1,19 414	0,48	0,35 553	+0,47 265	+0,59 398
0,04	0,39 862	+0,04 781	+1,19 204	0,49	0,35 381	+0,47 848	+0,57 213
0,05	0,39 844	+0,05 972	+1,18 936	0,50	0,35 207	+0,48 409	+0,55 010
0,06	0,39 822	+0,07 159	+1,18 608	0,51	0,35 029	+0,48 948	+0,52 791
0,07	0,39 797	+0,08 344	+1,18 221	0,52	0,34 849	+0,49 465	+0,50 556
0,08	0,39 767	+0,09 524	+1,17 775	0,53	0,34 667	+0,49 959	+0,48 308
0,09	0,39 733	+0,10 699	+1,17 271	0,54	0,34 482	+0,50 431	+0,46 048
0,10	0,39 695	+0,11 869	+1,16 708	0,55	0,34 294	+0,50 880	+0,43 777
0,11	0,39 654	+0,13 033	+1,16 088	0,56	0,34 105	+0,51 306	+0,41 497
0,12	0,39 608	+0,14 190	+1,15 410	0,57	0,33 912	+0,51 710	+0,39 208
0,13	0,39 559	+0,15 341	+1,14 676	0,58	0,33 718	+0,52 091	+0,36 913
0,14	0,39 505	+0,16 484	+1,13 885	0,59	0,33 521	+0,52 448	+0,34 613
0,15	0,39 448	+0,17 618	+1,13 038	0,60	0,33 322	+0,52 783	+0,32 309
0,16	0,39 387	+0,18 744	+1,12 137	0,61	0,33 121	+0,53 094	+0,30 003
0,17	0,39 322	+0,19 861	+1,11 180	0,62	0,32 918	+0,53 383	+0,27 696
0,18	0,39 253	+0,20 968	+1,10 170	0,63	0,32 713	+0,53 648	+0,25 390
0,19	0,39 181	+0,22 064	+1,09 106	0,64	0,32 506	+0,53 891	+0,23 085
0,20	0,39 104	+0,23 150	+1,07 990	0,65	0,32 297	+0,54 110	+0,20 783
0,21	0,39 024	+0,24 224	+1,06 823	0,66	0,32 086	+0,54 306	+0,18 486
0,22	0,38 940	+0,25 286	+1,05 604	0,67	0,31 874	+0,54 480	+0,16 195
0,23	0,38 853	+0,26 336	+1,04 335	0,68	0,31 659	+0,54 630	+0,13 912
0,24	0,38 762	+0,27 373	+1,03 018	0,69	0,31 443	+0,54 658	+0,11 636
0,25	0,38 667	+0,28 396	+1,01 651	0,70	0,31 225	+0,54 863	+0,09 371
0,26	0,38 768	+0,29 405	+1,00 238	0,71	0,31 006	+0,54 945	+0,07 116
0,27	0,38 466	+0,30 401	+0,98 777	0,72	0,30 785	+0,55 005	+0,04 874
0,28	0,38 361	+0,31 381	+0,97 273	0,73	0,30 563	+0,55 043	+0,02 646
0,29	0,38 251	+0,32 346	+0,95 723	0,74	0,30 339	+0,55 058	+0,00 433
0,30	0,38 139	+0,33 295	+0,94 130	0,75	0,30 114	+0,55 052	–0,01 764
0,31	0,38 023	+0,34 228	+0,92 495	0,76	0,29 887	+0,55 023	–0,03 944
0,32	0,37 903	+0,35 145	+0,90 819	0,77	0,29 659	+0,54 973	–0,06 106
0,33	0,37 780	+0,36 045	+0,89 103	0,78	0,29 430	+0,54 901	–0,08 248
0,34	0,37 654	+0,36 927	+0,87 348	0,79	0,29 200	+0,54 808	–0,10 369
0,35	0,37 524	+0,37 791	+0,85 555	0,80	0,28 969	+0,54 694	–0,12 468
0,36	0,37 391	+0,38 638	+0,83 726	0,81	0,28 737	+0,54 559	–0,14 545
0,37	0,37 255	+0,39 466	+0,81 862	0,82	0,28 504	+0,54 403	–0,16 597
0,38	0,37 115	+0,40 275	+0,79 963	0,83	0,28 269	+0,54 227	–0,18 624
0,39	0,36 973	+0,41 065	+0,78 032	0,84	0,28 034	+0,54 031	–0,20 626
0,40	0,36 827	+0,41 835	+0,76 070	0,85	0,27 798	+0,53 814	–0,22 600
0,41	0,36 678	+0,42 586	+0,74 077	0,86	0,27 562	+0,53 579	–0,24 546
0,42	0,36 526	+0,43 317	+0,72 056	0,87	0,27 324	+0,53 324	–0,26 464
0,43	0,36 371	+0,44 027	+0,70 007	0,88	0,27 085	+0,53 049	–0,28 351
0,44	0,36 213	+0,44 717	+0,67 932	0,89	0,26 848	+0,52 757	–0,30 202
0,90	0,26 609	+0,52 445	–0,32 034	1,38	0,15 395	+0,23 276	–0,73 890

Продолжение табл. 9

1	2	3	4	1	2	3	4
0,91	0,26 369	+0,52 116	–0,33 827	1,39	0,15 183	+0,22 537	–0,73 890
0,92	0,26 129	+0,51 769	–0,35 587	1,40	0,14 973	+0,21 800	–0,73 642
0,93	0,25 888	+0,51 404	–0,37 314	1,41	0,14 764	+0,21 065	–0,73 466
0,94	0,25 647	+0,51 023	–0,39 005	1,42	0,14 556	+0,20 331	–0,73 256
0,95	0,25 406	+0,50 624	–0,40 668	1,43	0,14 350	+0,19 600	–0,73 012
0,96	0,25 164	+0,50 210	–0,42 283	1,44	0,14 146	+0,18 871	–0,72 736
0,97	0,24 923	+0,49 779	–0,43 867	1,45	0,13 943	+0,18 145	–0,72 427
0,98	0,24 681	+0,49 332	–0,45 414	1,46	0,13 742	+0,17 423	–0,72 087
0,99	0,24 439	+0,48 871	–0,46 923	1,47	0,13 542	+0,16 704	–0,71 716
1,00	0,24 197	+0,48 394	–0,48 394	1,48	0,13 344	+0,15 988	–0,71 315
1,01	0,23 955	+0,47 903	–0,49 827	1,49	0,13 147	+0,15 277	–0,70 885
1,02	0,23 713	+0,47 398	–0,51 220	1,50	0,12 952	+0,14 571	–0,70 425
1,03	0,23 471	+0,46 879	–0,52 573	1,51	0,12 758	+0,13 869	–0,69 937
1,04	0,23 230	+0,46 346	–0,53 887	1,52	0,12 566	+0,13 172	–0,69 423
1,05	0,22 988	+0,45 801	–0,55 160	1,53	0,12 376	+0,12 481	–0,68 881
1,06	0,22 747	+0,45 243	–0,56 393	1,54	0,12 188	+0,11 795	–0,68 314
1,07	0,22 506	+0,44 675	–0,57 584	1,55	0,12 001	+0,11 114	–0,67 721
1,08	0,22 265	+0,44 092	–0,58 734	1,56	0,11 816	+0,10 440	–0,67 104
1,09	0,22 025	+0,43 499	–0,59 843	1,57	0,11 632	+0,10 977	–0,66 463
1,10	0,21 785	+0,42 895	–0,60 909	1,58	0,11 450	+0,09 111	–0,65 799
1,11	0,21 546	+0,42 281	–0,61 934	1,59	0,11 270	+0,08 456	–0,65 113
1,12	0,21 307	+0,41 657	–0,62 917	1,60	0,11 092	+0,07 809	–0,64 405
1,13	0,21 069	+0,41 023	–0,63 857	1,61	0,10 915	+0,07 168	–0,63 677
1,14	0,20 831	+0,40 380	–0,64 755	1,62	0,10 741	+0,06 535	–0,62 928
1,15	0,20 594	+0,39 728	–0,65 611	1,63	0,10 567	+0,05 910	–0,62 161
1,16	0,20 357	+0,39 067	–0,66 425	1,64	0,10 396	+0,05 292	–0,61 375
1,17	0,20 121	+0,38 399	–0,67 196	1,65	0,10 226	+0,04 682	–0,60 571
1,18	0,19 886	+0,37 724	–0,67 924	1,66	0,10 059	+0,04 081	–0,59 751
1,19	0,19 652	+0,37 041	–0,68 610	1,67	0,09 893	+0,03 487	–0,58 914
1,20	0,19 419	+0,36 352	–0,69 255	1,68	0,09 728	+0,02 903	–0,58 063
1,21	0,19 186	+0,35 656	–0,69 857	1,69	0,09 566	+0,02 326	–0,57 202
1,22	0,18 954	+0,34 955	–0,70 417	1,70	0,09 405	+0,01 759	–0,56 316
1,23	0,18 724	+0,34 248	–0,70 935	1,71	0,09 246	+0,01 200	–0,55 422
1,24	0,18 494	+0,33 536	–0,71 411	1,72	0,09 089	+0,00 650	–0,54 516
1,25	0,18 265	+0,82 820	–0,71 847	1,73	0,08 933	+0,00 110	–0,53 599
1,26	0,18 037	+0,32 099	–0,72 241	1,74	0,08 780	–0,00 422	–0,52 671
1,27	0,17 810	+0,31 375	–0,72 594	1,75	0,08 625	–0,00 944	–0,51 733
1,28	0,17 585	+0,30 648	–0,72 907	1,76	0,08 478	–0,01 456	–0,50 785
1,29	0,17 360	+0,29 917	–0,73 180	1,77	0,08 329	–0,01 959	–0,49 829
1,30	0,17 137	+0,29 984	–0,73 413	1,78	0,08 183	–0,02 453	–0,48 865
1,31	0,16 915	+0,28 449	–0,73 606	1,79	0,08 038	–0,02 937	–0,47 893
1,32	0,16 694	+0,27 712	–0,73 760	1,80	0,07 895	–0,03 411	–0,46 915
1,33	0,16 474	+0,26 974	–0,73 876	1,81	0,07 754	–0,03 875	–0,45 932
1,34	0,16 256	+0,26 235	–0,73 953	1,82	0,07 614	–0,04 329	–0,44 943
1,35	0,16 038	+0,25 495	–0,73 993	1,83	0,07 477	–0,04 774	–0,43 950
1,36	0,15 822	+0,24 755	–0,73 995	1,84	0,07 341	–0,05 208	–0,42 953
1,37	0,15 608	+0,24 016	–0,73 961	1,85	0,07 206	–0,05 633	–0,41 953
1,86	0,07 074	–0,06 047	–0,40 950	2,34	0,02 582	–0,14 955	+0,00 332

Продолжение табл. 9

1	2	3	4	1	2	3	4
1,87	0,06 943	–0,06 452	–0,39 946	2,35	0,02 522	–0,14 949	+0,00 915
1,88	0,06 814	–0,06 846	–0,38 940	2,36	0,02 463	–0,14 937	+0,01 485
1,89	0,06 687	–0,07 231	–0,37 934	2,37	0,02 406	–0,14 919	+0,02 040
1,90	0,06 562	–0,07 605	–0,36 928	2,38	0,02 349	–0,14 896	+0,02 582
1,91	0,06 438	–0,07 969	–0,35 923	2,39	0,02 294	–0,14 868	+0,03 109
1,92	0,06 316	–0,08 323	–0,34 918	2,40	0,02 239	–0,14 834	+0,03 623
1,93	0,06 195	–0,08 667	–0,33 916	2,41	0,02 184	–0,14 795	+0,04 122
1,94	0,06 077	–0,09 002	–0,32 916	2,42	0,02 134	–0,14 752	+0,04 608
1,95	0,05 959	–0,09 326	–0,31 919	2,43	0,02 083	–0,14 703	+0,05 079
1,96	0,05 844	–0,09 640	–0,30 925	2,44	0,02 033	–0,14 650	+0,05 537
1,97	0,05 730	–0,09 944	–0,29 936	2,45	0,01 984	–0,14 593	+0,05 981
1,98	0,05 618	–0,10 239	–0,28 950	2,46	0,01 986	–0,14 531	+0,06 411
1,99	0,05 508	–0,10 523	–0,27 970	2,47	0,01 888	–0,14 464	+0,06 828
2,00	0,05 399	–0,10 798	–0,26 996	2,48	0,01 842	–0,14 394	+0,07 231
2,01	0,05 292	–0,11 063	–0,26 027	2,49	0,01 797	–0,14 320	+0,07 620
2,02	0,05 186	–0,11 319	–0,25 064	2,50	0,01 753	–0,14 242	+0,07 997
2,03	0,05 082	–0,11 565	–0,24 109	2,51	0,01 709	–0,14 160	+0,08 360
2,04	0,04 980	–0,11 801	–0,23 160	2,52	0,01 667	–0,14 075	+0,08 710
2,05	0,04 879	–0,12 028	–0,22 220	2,53	0,01 625	–0,13 986	+0,09 047
2,06	0,04 780	–0,12 245	–0,21 287	2,54	0,01 585	–0,13 894	+0,09 375
2,07	0,04 682	–0,12 454	–0,20 363	2,55	0,01 545	–0,13 798	+0,09 683
2,08	0,04 586	–0,12 653	–0,19 448	2,56	0,01 506	–0,13 700	+0,09 982
2,09	0,04 491	–0,12 844	–0,18 542	2,57	0,01 468	–0,13 599	+0,10 298
2,10	0,04 398	–0,13 024	–0,17 646	2,58	0,01 431	–0,13 495	+0,10 542
2,11	0,04 307	–0,13 196	–0,16 759	2,59	0,01 394	–0,13 388	+0,10 804
2,12	0,04 217	–0,13 359	–0,15 883	2,60	0,01 358	–0,13 279	+0,11 053
2,13	0,04 128	–0,13 513	–0,15 017	2,61	0,01 323	–0,13 167	+0,11 291
2,14	0,04 041	–0,13 659	–0,14 162	2,62	0,01 289	–0,13 053	+0,11 517
2,15	0,03 955	–0,13 797	–0,13 318	2,63	0,01 256	–0,12 937	+0,11 732
2,16	0,03 871	–0,13 926	–0,12 486	2,64	0,01 223	–0,12 818	+0,11 935
2,17	0,03 788	–0,14 046	–0,11 665	2,65	0,01 191	–0,12 698	+0,12 127
2,18	0,03 706	–0,14 159	–0,10 856	2,66	0,01 160	–0,12 576	+0,12 308
2,19	0,03 626	–0,14 263	–0,15 920	2,67	0,01 130	–0,12 452	+0,12 479
2,20	0,03 547	–0,14 360	–0,09 274	2,68	0,01 100	–0,12 326	+0,12 638
2,21	0,03 470	–0,14 449	–0,08 502	2,69	0,01 071	–0,12 199	+0,12 787
2,22	0,03 394	–0,14 530	–0,07 743	2,70	0,01 042	–0,12 071	+0,12 926
2,23	0,03 319	–0,14 604	–0,06 996	2,71	0,01 014	–0,11 941	+0,13 055
2,24	0,03 246	–0,14 670	–0,06 263	2,72	0,00 987	–0,11 810	+0,13 174
2,25	0,03 174	–0,14 729	–0,05 542	2,73	0,00 961	–0,11 677	+0,13 283
2,26	0,03 103	–0,14 781	–0,04 835	2,74	0,00 935	–0,11 544	+0,13 387
2,27	0,03 034	–0,14 826	–0,04 141	2,75	0,00 909	–0,11 410	+0,13 473
2,28	0,02 965	–0,14 864	–0,03 461	2,76	0,00 885	–0,11 274	+0,13 555
2,29	0,02 898	–0,14 895	–0,02 794	2,77	0,00 861	–0,11 139	+0,13 427
2,30	0,02 833	–0,14 920	–0,02 141	2,78	0,00 837	–0,11 002	+0,13 691
2,31	0,02 768	–0,14 938	–0,11 502	2,79	0,00 814	–0,10 865	+0,13 746
2,32	0,02 705	–0,14 950	–0,00 877	2,80	0,00 792	–0,10 727	+0,13 793
2,33	0,02 643	–0,14 956	–0,00 265	2,81	0,00 770	–0,10 589	+0,13 832
2,82	0,00 748	–0,10 450	+0,13 863	3,30	0,00 172	–0,04 485	+0,09 690

1	2	3	4	1	2	3	4
2,83	0,00 727	–0,10 312	+0,13 886	3,31	0,00 167	–0,04 387	+0,09 549
2,84	0,00 707	–0,10 173	+0,13 902	3,32	0,00 161	–0,04 294	+0,09 409
2,85	0,00 687	–0,10 034	+0,13 910	3,33	0,00 156	–0,04 201	+0,09 268
2,86	0,00 668	–0,09 895	+0,13 912	3,34	0,00 151	–0,04 109	+0,09 128
2,87	0,00 649	–0,09 755	+0,13 906	3,35	0,00 146	–0,04 018	+0,08 987
2,88	0,00 631	–0,09 616	+0,13 894	3,36	0,00 141	–0,03 929	+0,08 847
2,89	0,00 613	–0,09 478	+0,13 875	3,37	0,00 136	–0,03 841	+0,08 707
2,90	0,05 595	–0,09 339	+0,13 850	3,38	0,00 132	–0,03 755	+0,08 567
2,91	0,00 578	–0,09 201	+0,13 819	3,39	0,00 127	–0,03 670	+0,08 428
2,92	0,00 562	–0,09 063	+0,13 782	3,40	0,00 123	–0,03 586	+0,08 290
2,93	0,00 545	–0,08 925	+0,13 739	3,41	0,00 119	–0,03 504	+0,08 151
2,94	0,00 530	–0,08 788	+0,13 691	3,42	0,00 115	–0,03 423	+0,08 014
2,95	0,00 514	–0,08 651	+0,13 638	3,43	0,00 111	–0,03 344	+0,07 877
2,96	0,00 499	–0,08 515	+0,13 579	3,44	0,00 107	–0,03 266	+0,07 741
2,97	0,00 485	–0,08 380	+0,13 515	3,45	0,00 104	–0,03 189	+0,07 606
2,98	0,00 470	–0,08 245	+0,13 446	3,46	0,00 100	–0,03 114	+0,07 471
2,99	0,00 457	–0,08 111	+0,13 373	3,47	0,00 097	–0,03 040	+0,07 338
3,00	0,00 443	–0,07 977	+0,13 296	3,48	0,00 094	–0,02 967	+0,07 205
3,01	0,00 430	–0,07 845	+0,13 214	3,49	0,00 090	–0,02 895	+0,07 074
3,02	0,00 417	–0,07 713	+0,13 128	3,50	0,00 087	–0,02 825	+0,06 943
3,03	0,00 405	–0,07 582	+0,13 038	3,51	0,00 084	–0,02 757	+0,06 814
3,04	0,00 393	–0,07 452	+0,12 944	3,52	0,00 081	–0,02 689	+0,06 685
3,05	0,00 381	–0,07 323	+0,12 847	3,53	0,00 079	–0,02 623	+0,06 558
3,06	0,00 370	–0,07 195	+0,12 747	3,54	0,00 076	–0,02 558	+0,06 432
3,07	0,00 358	–0,07 068	+0,12 643	3,55	0,00 073	–0,02 494	+0,06 308
3,08	0,00 348	–0,06 943	+0,12 536	3,56	0,00 071	–0,02 432	+0,06 184
3,09	0,00 337	–0,06 818	+0,12 426	3,57	0,00 068	–0,02 370	+0,06 062
3,10	0,00 327	–0,06 694	+0,12 313	3,58	0,00 066	–0,02 310	+0,05 941
3,11	0,00 317	–0,06 571	+0,12 198	3,59	0,00 063	–0,02 252	+0,05 821
3,12	0,00 307	–0,06 450	+0,12 080	3,60	0,00 061	–0,02 194	+0,05 703
3,13	0,00 298	–0,06 330	+0,11 960	3,61	0,00 059	–0,02 138	+0,05 586
3,14	0,00 288	–0,06 211	+0,11 838	3,62	0,00 057	–0,02 082	+0,05 471
3,15	0,00 279	–0,06 093	+0,11 714	3,63	0,00 055	–0,02 028	+0,05 357
3,16	0,00 271	–0,05 977	+0,11 588	3,64	0,00 053	–0,01 975	+0,05 244
3,17	0,00 262	–0,05 861	+0,11 460	3,65	0,00 051	–0,01 923	+0,05 133
3,18	0,00 254	–0,05 747	+0,11 330	3,66	0,00 049	–0,01 873	+0,05 023
3,19	0,00 246	–0,05 635	+0,11 199	3,67	0,00 049	–0,01 823	+0,04 915
3,20	0,00 238	–0,05 523	+0,11 066	3,68	0,00 046	–0,01 774	+0,04 808
3,21	0,00 231	–0,05 413	+0,10 933	3,69	0,00 044	–0,01 727	+0,04 703
3,22	0,00 224	–0,05 305	+0,10 798	3,70	0,00 042	–0,01 680	+0,04 599
3,23	0,00 216	–0,05 196	+0,10 662	3,71	0,00 041	–0,01 635	+0,04 497
3,24	0,00 210	–0,05 092	+0,10 525	3,72	0,00 039	–0,01 590	+0,04 396
3,25	0,00 203	–0,04 987	+0,10 387	3,73	0,00 038	–0,01 547	+0,04 297
3,26	0,00 196	–0,04 884	+0,10 249	3,74	0,00 037	–0,01 504	+0,04 200
3,27	0,00 190	–0,04 783	+0,10 110	3,75	0,00 035	–0,01 463	+0,04 103
3,28	0,00 184	–0,04 682	+0,09 970	3,76	0,00 034	–0,01 422	+0,04 009
3,29	0,00 178	–0,04 583	+0,09 830	3,77	0,00 033	–0,01 383	+0,03 916
3,78	0,00 031	–0,01 344	+0,03 824	3,88	0,00 021	–0,01 004	+0,02 991

Окончание табл. 9

1	2	3	4	1	2	3	4
3,79	0,00 030	–0,01 306	+0,03 734	3,90	0,00 020	–0,00 946	+0,02 842
3,80	0,00 029	–0,01 269	+0,03 646	3,92	0,00 018	–0,00 891	+0,02 699
3,81	0,00 028	–0,01 233	+0,03 559	3,94	0,00 017	–0,00 838	+0,02 562
3,82	0,00 027	–0,01 198	+0,03 473	3,96	0,00 016	–0,00 788	+0,02 430
3,84	0,00 025	–0,01 130	+0,03 307	3,98	0,00 014	–0,00 741	+0,02 303
3,86	0,00 023	–0,01 066	+0,03 146	4,00	0,00 013	–0,00 696	+0,02 181

1.3

ЛИТЕРАТУРА

1. Брукс, С. Применение статистических методов в метеорологии / С. Брукс, Н. Карузерс. – Л.: Гидрометиздат., 1963. – 416 с.

2. Булдык, Г. М. Теория вероятностей и математическая статистика: учеб. пособие для вузов / Г. М. Булдык. – Мн.: Выш. шк., 1989. – 285 с.

3. Вишнякова, С.В. Особенности яйцекладки рыжего соснового пилильщика в условиях Брянской области / С. В. Вишнякова // Совершенствование методов контроля лесопользования и состояния лесных экосистем. Научн. тр. – М.: МЛТИ, 1991. – Вып. 244. – С. 68-72.

4. Гайдук, В.Е. Экология размножения рыжей полевки в Беловежской пуще / В. Е. Гайдук, Е. С. Блоцкая // Заповедники Белоруссии. – Мн.: Ураджай, 1989. – Вып. 13. – С. 97-104.

5. Герасимович, А.И. Математическая статистика / А. И. Герасимович. – Мн.: Выш. шк., 1983. – 279 с.

6. Голубев, В. В. Элементы математической статистики в приложении к лесному делу / В. В. Голубев. – М.: Новая деревня, 1929. – 248 с.

7. Дворецкий, М. Л. Пособие по вариационной статистике / М. Л. Дворецкий. – 3-е изд. – М.: Лесная промышленность, 1971. – 104 с.

8. Длин, А. М. Математическая статистика в технике / А.М. Длин. – М.: Советская наука, 1958. – 465 с.

9. Дьячков, А. Н. Математическая статистика в применении к лесному делу / А. Н. Дьячков. – Архангельск: Архангельский лесотехнический институт, 1951. – 220 с.

10. Ефимова, М. Р. Общая теория статистики: учеб. пособие / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – 2-е изд. – М.: ИНФРА-М, 2000. – 413 с.

11. Коропачинский, И. Ю. Особенности естественного лесовозобновления лиственничных лесов нижнего пояса гор и предгорий хребта Танну-Ола / И. Ю. Коропачинский, В. С. Онучин // Леса Тувинской автономной области. Труды Сиб. технологического института. Сборник XXII. – Красноярск, 1959. – С. 53-68.

12. Лакин, Г. Ф. Биометрия: учеб. пособие для биол. спец. вузов / Г. Ф. Лакин. – 4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.

13. Луганский, Н. А. Влияние атмосферных промышленных загрязнений на семеношение и качество семян сосны / Н. А. Луганский, В. А. Калинин // Изв. высш. учеб. заведений. Лесн. журн. – 1990. – №1. – С. 7-10.

14. Математическая статистика: учебник / В. М. Иванова, В. Н. Калинина, Л. А. Нешумова и др. – 2-е изд., перераб. и доп. – М.: Высш. шк., 1981. – 371 с.

15. Машковский, В. П. Лесная биометрия: учеб.-метод. пособие по одноименной дисциплине для студентов специальности 1-75 01 01 «Лесное хозяйство» / В. П. Машковский. – Мн.: БГТУ, 2005. – 72 с.

16. Митропольский, А. К. Техника статистических вычислений / А. К. Митропольский. – М.: Наука, 1971. – 576 с.

17. Нестеров, В. Г. Новые методы повышения качества и продуктивности лесов / В. Г. Нестеров // Лесное хозяйство. – №11, 1952. – С. 28-35.

18. Плохинский, Н. А. Биометрия / Н. А. Плохинский. – М.: Издательство Московского университета, 1970. – 368 с.

19. Рокицкий, П. Ф. Биологическая статистика / П. Ф. Рокицкий. – Мн.: Вышэйшая школа, 1973. – 320 с.

20. Свалов, Н. Н. Вариационная статистика / Н. Н. Свалов. – М.: "Лесная промышленность", 1977. – 176 с.

21. Статистика: показатели и методы анализа: справ. пособ. / Н. Н. Бондаренко, Н. С. Бузыгина, Л. И. Василевская и др. – Мн.: Современная школа, 2005. – 628 с.

22. Трулль, О. А. Математическая статистика в лесном хозяйстве / О. А. Трулль. – Мн.: Высш. шк., 1966. – 234 с.

23. Тюрин, А. В. Основы вариационной статистики в применении к лесоводству / А. В. Тюрин. – М.: Гослесбумиздат, 1961. – 103 с.

24. Энтин, Л. И. Сосновый шелкопряд и его паразиты в условиях восточного Полесья / Л. И. Энтин // Лесохозяйственная наука и практика. – Мн.: Урожай, 1971. – Вып. 20. – С. 108-114.

СОДЕРЖАНИЕ